

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

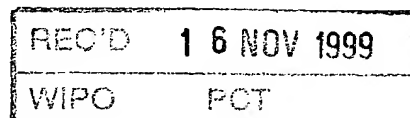
**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

Helsinki 07.10.99

09/830160
PCT/FI99/00870

ETUOIKEUSTODISTUS
PRIORITY DOCUMENT



Hakija
Applicant

GALILAEUS OY
Piispanristi

ETU

Patenttihakemus nro

982295

Patent application no

Tekemispäivä

23.10.98

Filing date

Kansainvälinen luokka
International class

C 12N

Keksinnön nimitys
Title of invention

**PRIORITY
DOCUMENT**
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH RULE 17.1(a) OR (b)

"Gene cluster involved in nogalamycin biosynthesis, and its
use in production of hybrid antibiotics"
(Nogalamysiinin biosynteesiin liittyvä geeniryhmittyä ja sen
käyttö hybridiantibioottien tuotossa)

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja
jäljennöksiä patentti- ja rekisterihallitukselle alkuaan
annetuista selityksestä, patenttivaatimuksista, tiivistelmästä
ja piirustuksista.

This is to certify that the annexed documents are true copies
of the description, claims, abstract and drawings originally
filed with the Finnish Patent Office.

Pirjo Kaila
Tutkimussihteeri

Maksu 300,- mk
Fee 300,- FIM

Osoite: Arkadiankatu 6 A
Address: P.O.Box 1160
FIN-00101 Helsinki, FINLAND

Puhelin: 09 6939 500
Telephone: + 358 9 6939 500

Telefax: 09 6939 5204
Telefax: + 358 9 6939 5204

Gene cluster involved in nogalamycin biosynthesis, and its use in production of hybrid antibiotics

Field of the invention

5

This invention relates to the gene cluster for nogalamycin biosynthesis derived from *Streptomyces nogalater*, and the use of the genes therein to obtain novel hybrid antibiotics for drug screening.

10 Background of the invention

Anthracyclines are antitumor antibiotics, mainly produced by *Streptomyces* sp. Daunomycin family of anthracyclines is commercially most important, since almost all of the around ten anthracyclines currently in clinical use, or in late clinical trials
15 for cytotoxic drugs, belong to this family. Despite the long history of anthracyclines, three decades or so, the studies on their biosynthesis are still going on, and there is further interest to obtain novel molecules for the development of cancer chemotherapy. A method currently used for finding novel molecules for drug screening is genetic engineering. Cloning the genes for anthracycline biosynthesis facilitates the
20 production of hybrid anthracyclines, as well as their use in combinatorial biosynthesis to generate novel molecules.

Nogalamycin, which was first described by Bhuyan and Dietz in 1965, is an anthracycline antibiotic produced by *Streptomyces nogalater*. It is highly active
25 against tumor cells, whereas toxic properties of this compound have prevented its progress to clinical trials (Bhuyan and Smith, 1975). However, menogaril (7-O-methylnogarol) is a semisynthetic derivative of nogalamycin, and its value in the treatment of cancer has been studied (e.g. Yoshida *et al.*, 1996), the interest being now mainly in Japan. Structurally nogalamycin (Fig. 1) differs from most other
30 anthracyclines, as e.g. from the daunomycin family, in two noteworthy features: (i) The stereochemistry at position nine is opposite, and (ii) it has a sugar moiety, in which nogalamine is attached at position 1 by a typical glycosidic bond, but it is also attached to carbon 2 by an extraordinary C-C bond. Structural elucidation of

nogalamycin was reported by Wiley *et al.* (1977). Furthermore, biosynthetic studies of nogalamycin have been published by Wiley *et al.* in 1978 giving information of the building blocks: The aglycone moiety is built from ten acetates; the neutral sugar, nogalose, is derived from glucose; and methyl groups of both of the sugars, nogalamine and nogalose, are transferred from methionine. The origin of nogalamine was not clearly solved by Wiley, but most probably nogalamine is also derived from glucose.

Molecular cloning of biosynthesis genes for anthracyclines has facilitated the studies on molecular genetics, providing tools for rational modifications of the structures, while also for surprising combinations with other antibiotics. Most of the interest has focused on daunomycin biosynthesis genes, as reported in several publications (Lomovskaya *et al.*, 1998; Rajgarhia and Strohl, 1997 and references therein). Some genes for aclacinomycin biosynthesis from *S. galilaeus* (Fujii and Ebizuka, 1997) and for rhodomycin biosynthesis from *S. purpurascens* (Niemi *et al.*, 1994) have been cloned as well. We have cloned the biosynthesis genes for nogalamycin, and successfully used the genes for producing hybrid anthracyclines. Most of the genes are involved in polyketide pathway, being responsible for the formation of a tricyclic intermediate, and they are reported in Ylihonko *et al.*, 1996a and b, and by Torkkell *et al.*, 1997. Despite the advances in molecular cloning, the biosynthetic pathway from glucose to sugars found in anthracyclines is still mainly hypothetical.

Regarding the genes for deoxyhexose pathway, Madduri *et al.* (1998) have reported that a gene derived from avermectin biosynthesis cluster caused the production of hybrid anthracyclines altering the sugar moiety when transferred into an *S. peucetius* mutant. The product obtained was epirubicin, a commercially important anthracycline. In this case a hydroxy group in the daunosamine moiety was in the opposite stereochemistry due to the action of an avermectin biosynthesis gene. *S. galilaeus* has been used as the host to prepare hybrid anthracyclines using the genes derived from rhodomycin pathway from *S. purpurascens* (Niemi *et al.*, 1994), and from nogalamycin biosynthesis cluster from *S. nogalater* (Ylihonko *et al.*, 1996a). The genes for nogalamycin pathway were used to generate the hybrid anthracycline production in *S.*

steffisburgensis producing typically steffimycin (Kunnari *et al.*, 1997). Previously, biosynthesis genes for actinorhodin have been expressed in *S. galilaeus* resulting in the formation of aloesaponarin (Strohl *et al.*, 1991). These hybrid compounds were modified in the aglycone moiety.

5

Summary of the invention

The present invention concerns a gene cluster of *Streptomyces nogalater*, most of the genes of which are derived from the deoxyhexose pathway for nogalamine and
 10 nogalose. Expressing a DNA fragment of the said region in *S. galilaeus*, which produces aclacinomycins, hybrid anthracyclines are obtained, in which the aglycone moiety is derived from *S. galilaeus*, whereas the sugar moiety is characteristic neither to *S. nogalater* nor to *S. galilaeus*. Furthermore, when inserting the gene included in said cluster, encoding a cyclase for nogalamycin, into a suitable plasmid construction,
 15 nogalamycinone is obtained, which is the aglycone of nogalamycin. Since the stereochemistry of nogalamycin differs from most other anthracyclines, using this gene enables the preparation of C-9 stereoisomers of the anthracycline molecules.

Detailed description of the invention

20

The experimental procedures of the present invention are methods conventional in the art. The techniques not described in detail here are given in the manuals by Hopwood *et al.* "Genetic manipulation of Streptomyces: a laboratory manual" The John Innes Foundation, Norwich (1985) and by Sambrook *et al.* (1989) "Molecular cloning: a
 25 laboratory manual". The publications, patents and patent applications cited herein are given in the reference list in their entirety.

30

The present invention concerns particularly the gene cluster for nogalamycin biosynthesis (*Sno5*-cluster) causing the production of hybrid antibiotics with
 30 modifications in the sugar moiety. The invention concerns in specific the use of the genes for nogalamine/nogalose biosynthesis to generate hybrid antibiotics modified in sugar moieties. The invention also concerns the use of a specific cyclase gene

included in the gene cluster of the invention, to generate the C-9 stereoisomers of typical anthracyclines.

5 The gene cluster according to the present invention is linked to the earlier reported clusters for nogalamycin biosynthesis. The starting point of the present invention was the gene cluster for nogalamycin chromophor (International Patent Application WO 96/10581). ~~Subsequently, we have found some genes for the deoxyhexose pathway of~~
~~nogalamycin biosynthesis (Torkkell *et al.*, 1997), and a part of the fragment~~
~~comprising said genes was used to clone the genes for this invention.~~

10

The biosynthesis genes for nogalamycin can be isolated from *Streptomyces* sp., particularly from *S. nogalater*, which produces nogalamycin. Species which produce nogalamycin-like anthracyclines can also be used, e.g. *S. violaceochromogenes* producing arugomycin (Kawai *et al.*, 1987), or *S. avidinii* producing avidinorubicin
15 (Aoki *et al.*, 1991).

Genomic DNA of a *Streptomyces* strain carrying the genes for nogalamycin

biosynthesis is used in preparing a genomic library. Suitable gene fragments for cloning may be obtained by any frequently digesting restriction enzyme. Typically
20 *Sau*3AI is used. The isolated fragments could be inserted by ligation in any *Escherichia coli* vector such as a plasmid, a phagemid, a phage, or a cosmid. A cosmid vector is preferred since it enables the cloning of large DNA fragments. A cosmid vector such as pFD666 (ATCC No. 77286) is suitable for this purpose, as it enables cloning of the fragments of about 40 kb. The *Bam*HI site of pFD666, giving
25 sticky ends to the *Sau*3AI fragments may be used for cloning. Commercially available kits may be used to pack the DNA in phage particles. Various *E. coli* strains can be used for the infection by the DNA packed. An appropriate *E. coli* strain is, e.g. XL1Blue MRF⁺, which is deficient in several restriction systems.

30 Using *E. coli* as a host strain for the genomic library, hybridization is an advantageous screening strategy. The probe for hybridization may be any known fragment derived from the nogalamycin gene cluster, but a short fragment of about 1

kb derived from one end of the biosynthetic region previously cloned is preferred. Colonies for the genomic library are transferred for filter hybridization to membranes, preferably to nylon membranes. Since the average size for a genomic DNA fragment is 40 kb, 2300 colonies gave 99,99% probability to find the expanded region for nogalamycin biosynthesis. Any method for hybridization may be used but, in particular, the DIG System (Boehringer Mannheim, GmbH, Germany) is useful. Since the probe is homologous to the hybridized DNA, it is preferable to carry out the stringent washes of hybridization at 70°C in a low salt concentration according to Boehringer Mannheim's manual "DIG System User's Guide for Filter Hybridiza-
 tion". At least 80% homology is suggested to be needed for a DNA fragment to bind a probe in the conditions used for washes.

Using this protocol, seven clones out of about 5000 gave positive signals, and were picked up for DNA isolation. Restriction mapping is an appropriate technique for characterizing the clones. The positive clones may be digested with convenient restriction enzymes to demonstrate the physical linkage map of the DNA fragments. The cosmid used for cloning was a shuttle cosmid replicating in both *E. coli* and *Streptomyces* sp. However, the transfer of the recombinant cosmids in *S. lividans* TK24, which is a typically used laboratory strain in cloning *Streptomyces*, resulted in deletions, and was omitted. Instead, we rather used in the expression studies the plasmid pIJ486, a high copy number *Streptomyces* plasmid. However, any plasmid being able to stably replicate in *Streptomyces* may be used for this purpose.

Two *Bgl*II fragments of one of the clones were separately inserted into pIJ486 vectors, and the two plasmids obtained were transferred into a primary host, *S. lividans* TK24. The recombinant plasmids obtained (pSY42 and pSY43), containing a 10 kb and a 7 kb fragment from *S. nogalater* genomic DNA, respectively, were isolated from the primary host and further introduced into other *Streptomyces* strains by protoplast transformation. The recombinant plasmid containing the 10 kb fragment caused the production of hybrid anthracyclines in the *S. galilaeus* mutant strain H039, which endogenously produces aklavinone-rhodinose-rhodinose-rhodinose. A few other *S. galilaeus* strains (H075, H026, H063) mutated in deoxyhexose pathway

for sugars in aclacinomycin were used in transformation, and new hybrid compounds were obtained. Since the structure of nogalamycin is almost unique among anthracyclines, the plasmids could be transferred to other anthracycline-producing strains, such as *S. peucetius*, which produces daunomycin, and *S. purpurascens*, which produces rhodomycins, to modify the structures of the characteristic antibiotics.

As the cloned cluster was linked to nogalamycin biosynthesis region already known, its ability to generate the modification in sugar moiety suggested the presence of the genes for deoxyhexose pathway. However, sequencing is necessary to deduce the function of the genes in the cluster cloned. The DNA fragments of 10 kb and 7 kb were further inserted into the plasmid pSL1190 for subcloning. Sequencing strategies such as a deletion set of the DNA fragments, shotgun cloning or primer walking could be used, but we prefer to use restriction fragments for subcloning. Using ABI PRISM system (Perkin-Elmer) for sequencing it is possible to get 500 to 700 bases per one reaction, which means that about 1 kb fragments sharing overlapping bases are needed for sequencing. For this purpose, 27 subclones were constructed.

Sequencing of the flanked *Bgl*II fragments consisting of about 16000 bp revealed 15 complete ORFs. The sequence analysis can be made by any computer based program, such as GCG (Madison, Wisconsin, USA) package. According to the present invention the putative gene functions as deduced from the sequence homology of those available in the libraries are

- aminotransferase (*snogI*), not completed
- 1. dTDP-glucose synthase (*snogJ*)
- 2. aminomethyl transferase (*snogA*)
- 3. polyketide cyclase, (*snoaM*)
- 4. a gene of deoxyhexose pathway, unknown (*snogN*)
- 5. hydroxylase, (*snoaG*)
- 6. dTDP-4-dehydrorhamnose reductase (*snogC*)
- 7. dTDP-glucose 4,6-dehydratase (*snogK*)
- 8. NAME cyclase (*snoaL*)

9. unknown (*snoK*)
 10. glycosyl transferase, GTF (*snogD*)
 11. unknown (*snoW*)
 12. glycosyl transferase, GTF (*snogE*)
 - 5 13. unknown (*snoL*)
 14. unknown (*snoO*)
 15. C-7 ketoreductase (*snoaF*)
-
- unknown (*snoN*), not completed
-

- 10 Gene designations: g means that the gene involved in biosynthesis of the glycosidic proportion including glycosyl transferases, whereas a points out that the gene is needed for the formation of the aglycone moiety.

- 15 Considering the proposed biosynthetic pathway for nogalamycin shown in Fig 3. we are able to cause several changes for the structures of antibiotics by the genes identified, including *snoaL*, responsible for the cyclization of the fourth ring of the aglycone moiety while determining the stereochemistry of the anthracyclinone, and the genes affecting the formation of nogalamine and nogalose (*snogJ*, *snogK*, *snogN*, *snogC*, *snogA*), and, in addition, the genes responsible for joining the sugar residues to the aglycone moiety (*snogD* and *snogE*).

- 20 These genes could be separately inserted in a vector using suitable restriction sites, or by amplifying the genes by PCR. The fragments may contain an intrinsic promoter, or a promoter may be separately cloned. It is advantageous to use a vector carrying a promoter to allow expression of the genes in a *Streptomyces* strain. The plasmid
- 25 pIJE486 contains a promoter *ermE* for erythromycin resistance gene, allowing constitutive expression of the genes inserted in a correct orientation. Special attention is drawn to the gene encoding a cyclase for the aliphatic ring, but any gene of said cluster may be expressed in *Streptomyces* hosts. The said cyclase converts the
- 30 stereochemistry at C9 of auramycinone in TK24, if inserted into the plasmid possessing the other genes for auramycinone biosynthesis, except the cyclase responsible for the typical stereochemistry of anthracyclines.

Streptomyces strains, in particular *S. galilaeus*, carrying the recombinant plasmids are cultivated in media wherein antibiotics are produced. The hybrid compounds are extracted with organic solvents from the culture broth, and the compounds are separated and purified using chromatographic techniques.

5

According to this invention *S. galilaeus* H039 carrying the plasmid pSY42 and designated as H039/pSY42 produces aklavinone-4'-epi-2-deoxyfucose in E1 medium supplemented with thiostrepton to give selection pressure for the plasmid containing strains.

10

S. lividans TK24 carrying the plasmid pSY15c containing the genes for the nogalamycin chromophore and the genes for a cyclase (*snoaL*) and a ketoreductase (*snoaF*), was cultivated in E1 medium supplemented with thiostrepton. The compound 9-epi-auramycinone was produced, and this structure is now called nogalamycinone. Any DNA fragment of the invention subcloned from a 17 kb nogalamycin biosynthesis region can be inserted in a vector replicating in *Streptomyces*, and the products may be produced by fermentation of the plasmid containing strains.

15

20 Brief description of the drawings

Fig. 1 shows the structures of nogalamycin, daunomycin and aclacinomycin.

Fig. 2 is a diagram of the gene cluster (*Sno5*) of the invention for nogalamycin biosynthesis.

25

Fig. 3 describes the proposed biosynthesis pathway for nogalamycin.

Fig. 4 shows a diagram of the plasmid pSY15c. The genes *snoaL* (aL) and *snoaF* (aF) shown black are inserted in the plasmid pSY15 to give pSY15c. aL represents a cyclase *snoaL* and aF is for C-7 ketoreductase *snoaF*. pSY15 (WO 96/10581) generates the production of a tricyclic intermediate for

30

nogalamycin biosynthesis in *S. lividans*. The abbreviations **a1**, **a2** and **a3** refer to the genes *snoa1*, *snoa2* and *snoa3*, respectively, for minimal PKS. **rA** is the *snoA* gene for an activator, **aB** is the *snoaB* gene for oxygenase, **aC** is the *snoaC* gene for methylase, **aD** is the *snoaD* gene for polyketide ketoreductase and **aE** is the *snoaE* for aromatase. **gF** (the *snogF* gene) and **gG** (the *snogG* gene) involved in the deoxyhexose pathway are not functional in the construct. ~~**aph** is an aminoglycoside phosphotransferase gene,~~ and **tsr** is a thiostreptone resistance gene.

10 Examples to further illustrate the invention are given hereafter.

EXPERIMENTAL

Materials used

15 Restriction enzymes used were purchased from Promega (Madison, Wisconsin, USA) or Boehringer Mannheim (Germany), and alkaline phosphatase from Boehringer Mannheim, and used according to the manufacturers' instructions. Proteinase K was purchased from Promega and lysozyme from Sigma (St. Louis, USA). HybondTM-N nylon membranes used in hybridization were purchased from Amersham
20 (Buckinghamshire, England), DIG DNA Labelling Kit and DIG Luminescent Detection Kit from Boehringer Mannheim. Qiaquick Gel Extraction Kit from Qiagen (Hilden, Germany) was used for isolating DNA from agarose.

Bacterial strains and their use

25 *Escherichia coli* XL1 Blue MRF' was used for cloning.
Streptomyces nogalater ATCC 27451; the gene cluster of nogalamycin biosynthesis was cloned from this strain.
The host strains to express the genes cloned were:
Streptomyces lividans TK24, also used as a primary host to clone DNA propagated in
30 *E. coli*.
Streptomyces galilaeus H039, produces aklavinone-rhodinose-rhodinose-rhodinose

Streptomyces galilaeus H026, produces aclacinomycin N, ACMN, (aklavinone-rhodosamine-2-deoxyfucose-rhodinose)

Streptomyces galilaeus H063, produces aklavinone

Streptomyces galilaeus H075, produces aklavinone-rhodosamine-2-deoxyfucose-2-deoxyfucose

The detailed description of the mutants H039 and H026 is given in Ylihonko *et al.* (1994) and of H075 in the FI patent application No. 981062 (Ylihonko *et al.*, 1998). H063 has not been described in the literature but it was obtained by NTG

mutagenesis of *S. galilaeus*, and selected to be used as the host strain in the hybrid compound production, as it accumulates aklavinone without any sugar residues.

Plasmids

E. coli - *Streptomyces* shuttle cosmid pFD666 (ATCC 77286) was used for cloning the chromosomal DNA. *E. coli* cloning vectors pSL1190 (Pharmacia) and pUC19 were used for preparing the subclones.

pIJ486 is a high copy plasmid vector provided by prof. Sir David Hopwood, John Innes Centre, UK (Ward *et al.*, 1986)

pIJE486 is a vector containing *ermE* gene in the polylinker of pIJ486 (Bibb *et al.*, 1985).

pSY15 is a pIJ486 based plasmid construct, wherein the genes of polyketide pathway for nogalamycin biosynthesis were cloned (Ylihonko *et al.*, 1996a).

Nutrient media and solutions

For cultivation of *S. nogalater* for total DNA isolation TSB medium was used.

Lysozyme solution (0.3 M sucrose, 25 mM Tris, pH 8 and 25 mM EDTA pH 8) was used in isolation of total DNA. TE buffer (10 mM Tris, pH 8.0 and 1mM EDTA) was used to dissolve the DNA.

TRYPTONE-SOYA BROTH (TSB)

Per litre: Oxoid Tryptone Soya Broth powder 30 g.

ISP4

5 Bacto ISP-medium 4, Difco; 37 g/l.

E1 Per litre in tap water:

| | | |
|----|--|-------|
| | glucose | 20 g |
| | soluble starch | 20 g |
| 10 | Farmamedia | 5 g |
| | yeast extract | 2.5 g |
| | K ₂ HPO ₄ •3H ₂ O | 1.3 g |
| | MgSO ₄ •7H ₂ O | 1 g |
| | NaCl | 3 g |
| 15 | CaCO ₃ | 3 g |

pH adjusted to 7.4 before autoclaving

20 General methods

NMR data was collected with a JEOL JNM-GX 400 spectrometer at the ambient temperature. ¹H and ¹³C NMR samples were internally referenced to TMS.

25 The anthracycline metabolites were detected by HPLC (LaChrom, Merck Hitachi, pump L-7100, detector L-7400 and integrator D-7500) using a LiChroCART RP-18 column (4.6x250mm). Acetonitrile:potassium hydrogen phosphate buffer (60 mM, pH 3.0 adjusted with citric acid) was used as the mobile phase. Gradient system starting from 65% to 30% of potassium dihydrogen phosphate buffer was used to separate the compounds. The flow rate was 1 ml/min and the detection was effected at 430 nm.

30

ISP4 plates supplemented with thiostrepton (50 µg/ml) were used to maintain the plasmid carrying cultures.

Example 1. Cloning the gene cluster for nogalamycin biosynthesis

1.1 Cosmid library

For the isolation of total DNA, *Streptomyces nogalater* (ATCC 27451) was grown
 5 for three days in 50 ml of TSB medium supplemented with 0.5% of glycine. The
 cells were harvested by centrifuging for 15 min at 3900 x g in 12 ml Falcon tubes,
 and the cells were stored at -20°C. Cells from a 12 ml sample of the culture were
 used to isolate the DNA. 5 ml of lysozyme solution containing 5 mg of lysozyme/ml
 was added onto the cells, incubated for 20 min at 37°C. 500 µl of 10% SDS
 10 containing 0.7 mg of proteinase K was added onto the cells and incubated for 80 min
 at 62°C, another 500 µl of 10% SDS containing 0.7 mg of proteinase K was added,
 and incubation was continued for 60 min. The sample was chilled on ice and 600 µl
 of 3M NaAc, pH 5.8 were added, and the mixture was extracted with equilibrated
 phenol (Sigma). The phases were separated by centrifuging at 1400 x g for 10 min.
 15 The DNA was precipitated from the water-phase with equal volume of isopropanol to
 spool with a glass rod, and washed by dipping to 70% ethanol, air dried and
 dissolved in 500 µl of TE-buffer.

The chromosomal DNA was partially digested with *Sau3AI*. The DNA fragments
 20 were separated by agarose gel electrophoresis, and the fragments of 30 to 50 kb were
 cut from the 0.3% low gelling temperature SeaPlaque® agarose. The DNA bands
 were isolated from the gel by heating to 65°C, extracting with equal volume of
 equilibrated phenol, and the phases were separated by centrifuging for 15 min at
 2500 x g. The phenol phase was extracted with TE buffer, centrifuged and the water
 25 phases were pooled. The DNA was precipitated by adding 0.1 volumes of NaAc, pH
 5.8 and 2 volumes of ethanol at -20°C for 30 min, centrifuged for 30 min at 15 000
 rpm in Sorvall RC5C centrifuge using SS-34 rotor with adapters for 10 ml tubes.
 The pellet was air dried and dissolved in 20 µl of TE buffer. The isolated fragments
 were ligated to pFD666 cosmid vector digested with *Bam*HI and dephosphorylated.
 30 The DNA was packed to phage particles, and infected to *E. coli* using Gigapack® III
 XL Packing Extract Kit according to the manufacturer's instructions.

1.2 Identification of the clones by hybridization

- The infected cells were grown on LB plates containing 50 $\mu\text{g/ml}$ kanamycin and transferred to HybondTM-N nylon membranes (Amersham). The membranes were handled according to the protocol described in Boehringer Mannheim's manual "The DIG System User's Guide for Filter Hybridization". The probe used to screen the colonies for an expanded nogalamycin gene cluster was a 1.07 kb *SacI* fragment from the cluster described earlier (Torkkell *et al.*, 1997). The plasmid carrying the probe was digested with *SacI*, and the fragment was separated from the vector by agarose gel electrophoresis and isolated from the gel using Qiaquick Gel Extraction Kit (Qiagen). The probe was labelled by digoxigenin using random prime labelling system according to Boehringer Mannheim's manual "The DIG System User's Guide for Filter Hybridization". 5000 colonies were screened by hybridization at 70°C using the probe described. Positive colonies were detected using DIG Luminescent Detection Kit (Boehringer Mannheim). Seven colonies gave a positive signal.
- Cosmids from the positive clones were isolated from a 5ml culture by alkaline lysis method. Restriction analysis showed that the cloned fragments overlapped each other representing at least 60 kb of the continuous DNA. The positive clones obtained were designated as pFDSno1 to pFDSno7.

1.3. Subcloning the fragments for sequencing

- Clone No. 5, designated as pFDSno5, was digested with *BglII*, and for subcloning two fragments of about 10 kb and 7 kb were isolated and ligated to pSL1190 digested with *BglII* and dephosphorylated. The plasmids obtained were named as pSn42 and pSn43, respectively. These two fragments cover the DNA region flanked to the previously characterized area of nogalamycin biosynthesis cluster. To determine the nucleotide sequence of the whole 17 kb region cloned in pSn42 and pSn43 the convenient restriction sites were used to subclone the fragments to the vector pUC19 or pSL1190 giving 16 subclones from the insert of pSn42 and 11 subclones of pSn43.
- E. coli* XL1 Blue MRF' cells were cultivated overnight at 37 °C in 5 ml of LB-medium supplemented with 50 $\mu\text{g/ml}$ of ampicillin. To isolate plasmids for

sequencing reactions Wizard Plus Minipreps DNA Purification System kit of Promega, or Biometra silica spin plasmid miniprep kit of Biomedizinische Analytik GmbH were used according to the manufacturers' instructions.

- 5 DNA sequencing was performed using the automatic ABI DNA sequenator (Perkin-Elmer) according to the manufacturer's instructions.

1.4 Sequence analysis and the deduced functions of the genes

~~Sequence analyses were effected using the GCG sequence analysis software package~~

- 10 (Version 8; Genetics Computer Group, Madison, Wisconsin, USA). The translation table was modified to accept also GTG as a start codon. Codon usage was analysed using published data (Wright and Bibb 1992).

- 15 According to the CODONPREFERENCE program the sequenced DNA fragment contained 15 complete open reading frames (ORFs), and the 5' end of other two ORFs in the both ends of the fragment according to the invention. The functions of the genes were concluded by comparing the amino acid sequences translated from their base sequences to the known protein sequences in the data banks. The results are shown in Table 1. The positions given refer to the appended sequence listing.
- 20 The amino acid sequences of the peptides are given in SEQ ID NO:2 to SEQ ID NO:18.

Table 1

| Gene | Position | Amino acids (SEQ ID NO) | Deduced function | Remarks |
|--------------|----------------------|----------------------------|--|---------|
| <i>snogI</i> | -1027 compl | >342 (2) | aminotransferase | 5' end |
| <i>snogJ</i> | 1192-2073 | 293 (3) | dTDP-glucose synthase | |
| <i>snogA</i> | 2106-2822 compl | 238 (4) | aminomethyl transferase | |
| <i>snoaM</i> | 2826-3800 compl | 324 (5) | a polyketide cyclase | |
| <i>snogN</i> | 3799-5025 | 408 (6) | <i>dnrQ</i> homology (Otten <i>et al.</i> , 1995), unknown | |
| <i>snoaG</i> | 5088-6356 | 422 (7) | hydroxylase | |
| <i>snogC</i> | 6334-7209 compl | 291 (8) | dTDP-4-dehydrorhamnose reductase | |
| <i>snogK</i> | 7245-8297 compl | 350 (9) | dTDP-glucose-4,6-dehydratase | |
| <i>snoaL</i> | 8537-8941 | 134 (10) | NAME cyclase (nogalonic acid methyl-ester) | |
| <i>snoK</i> | 8992-9699 | 235 (11) | unknown | |
| <i>snogD</i> | 9745-10917 compl | 390 (12) | glycosyl transferase | |
| <i>snoW</i> | 11057-11884 | 275 (13) | unknown | |
| <i>snogE</i> | 11928-* | >424 (14) | glycosyl transferase | |
| <i>snoL</i> | 13335-13754 compl | 139 (15) | unknown | |
| <i>snoO</i> | 13974-14441 | 155 (16) | homologous to <i>mtmX</i> of mithramycin cluster | |
| <i>snoaF</i> | 14532-15377 | 281 (17) | C-7 ketoreductase, analogous to aklaviketone ketoreductase | |
| <i>snoN</i> | 15450-* | >190 (18) | unknown | 5' end |

*, nucleotide sequence of about 100 bp, not known

1.5 Expression cloning

The 10 kb *Bgl*III fragment from pFDSno5 was cloned into the plasmid pIJ486 and the plasmid obtained was named as pSY42. Correspondingly, the 7 kb *Bgl*III fragment from pFDSno5 was cloned into the plasmid pIJE486, and the plasmid pSY43 was
 5 obtained. Plasmid pSY42 was introduced into *S. lividans* strain TK24 by protoplast transformation, isolated from it and further introduced into *S. galilaeus* mutant H039, and after propagation in H039, transferred to other *S. galilaeus* mutants blocked in the deoxyhexose pathway for characteristic sugars of aclacinomycins (H075, H026, and H063). E1 medium was used for anthracycline production, and the products were
 10 extracted from the culture with toluene:methanol (1:1) at pH 7. Anthracycline metabolites were analyzed by HPLC. The products of the mutants H039, H026, H063 and H075 carrying pSY42 differed from those obtained by the mutants without the plasmid.

15 According to the sequence analysis pSY42 contained a cyclase designated as NAMEC (nogalonic acid methyl ester cyclase), and in pSY43 a ketoreductase gene was identified. Expression constructions were prepared which contained all the genes needed for the formation of nogalamycin aglycone. A 1.4 kb *Bam*HI-*Sac*I fragment from pSY42 (containing NAMEC) and a 1.1 kb *Mlu*I-*Kpn*I fragment from pSY43
 20 carrying the gene for a ketoreductase of C-7 keto group were ligated to pSY15 linearized by *Sac*I, to form the plasmid pSY15c (Fig. 4). Plasmid pSY15c was introduced into *S. lividans* TK24, and the strain TK24/pSY15c was cultivated in E1 medium supplemented with thiostrepton. An aglycone compound was produced, and this structure is now called nogalamycinone.

25

Example 2. Compounds generated by the *sno5*-cluster

2.1 Production and purification of the products derived from H039/pSY42 and TK24/pSY15c

30 The seed culture, 180 ml of E1 culture of the plasmid containing strain, H039/pSY42 or TK24/pSY15c, was obtained by cultivating the strain in three 250 ml Erlenmeyer flasks containing 60 ml of E1 medium supplemented with thiostrepton (5 µg/ml) for

four days at 30°C, 330 rpm. The combined culture broths (180 ml) were used to inoculate 13 l of E1 medium in a fermentor (Biostat E). Fermentation was carried out for seven days at 28°C (330 rpm, aeration: 450 l/min).

- 5 The cells were harvested by centrifuging. 2.6 l of methanol was used to break the bacterial cells and to extract anthracycline metabolites accumulated. The anthracycline metabolites were extracted using 2 l of dichloromethane at pH 6. The organic layer was evaporated to dryness. The viscous residue was flashed through a polyamide (11) column using water:methanol from 1:9 to 0:10 as the eluent. Pooled
- 10 fractions containing the compounds were further purified on a Merck-Hitachi HPLC using preparative reversed phase column (LichroCART RP-18, 5 μ m) with mobile phase acetonitrile:1 % AcOH in water (1:1). Evaporation of acetonitrile gave pure products as yellow powders dried under vacuum.

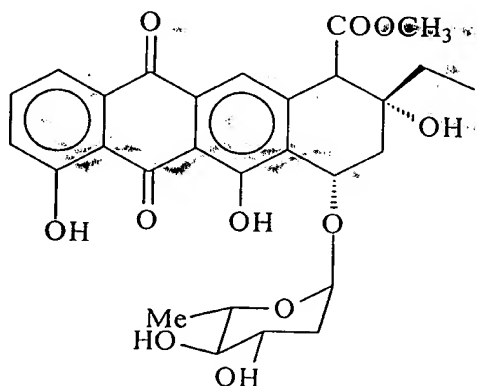
15 2.2 Structural elucidation of the compounds derived from H039/pSY42 and from TK24/pSY15c

NMR analysis included ¹H NMR, ¹³C NMR, NOE, DEPT and HMBG techniques. Protons were assigned using NOESY and 2D pTOCSY techniques and carbons using DEPT and HMBG techniques.

20

As deduced from the data given in Tables 2 and 3, the structures revealed were aklavinone-4'-epi-2-deoxyfucose from the culture of H039/pSY42, and 9-epi-auramycinone (=nogalamycinone) from the culture of TK24/pSY15c. The chemical structures of the compounds are shown below in Formula I and Formula II, respectively.

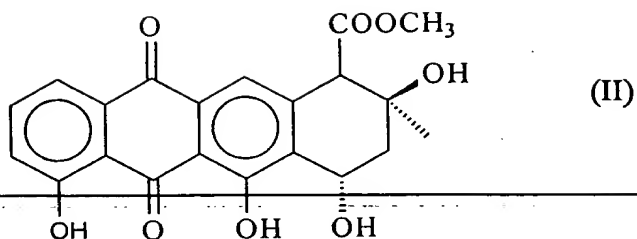
25



(I)

30

5



10

Deposited microorganisms

The following microorganisms were deposited according to the Budapest Treaty at Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1b, D-38124 Braunschweig, Germany.

15

| Microorganism | Accession number | Date of deposit |
|---------------|------------------|-----------------|
|---------------|------------------|-----------------|

| | | |
|-------------------------------|--|--|
| <i>S. lividans</i> TK24/pSY42 | | |
|-------------------------------|--|--|

20

| | | |
|----------------------------|--|--|
| carrying the plasmid pSY42 | | |
|----------------------------|--|--|

| | | |
|--|-----------|--|
| | DSM 12451 | |
|--|-----------|--|

| | | |
|--|--|-----------------|
| | | 14 October 1998 |
|--|--|-----------------|

| | | |
|-------------------------------|--|--|
| <i>S. lividans</i> TK24/pSY43 | | |
|-------------------------------|--|--|

| | | |
|----------------------------|--|--|
| carrying the plasmid pSY43 | | |
|----------------------------|--|--|

| | | |
|--|-----------|--|
| | DSM 12452 | |
|--|-----------|--|

| | | |
|--|--|-----------------|
| | | 14 October 1998 |
|--|--|-----------------|

Table 2. ^1H and ^{13}C assignments of the compound aklavinone-4'-epi-2-deoxyfucose (Formula I).

| Site | ^1H | ^{13}C |
|------|--------------------------------|-----------------|
| 1 | 7.74, 1H, dd, 7.5, 1.3 | 120.1 |
| 2 | 7.68, 1H, dd, 8.4, 7.5 | 137.3 |
| 3 | 7.27, 1H, dd, 8.3, 1.3 | 124.6 |
| 4 | — | 161.9 |
| 4-OH | 11.70, 1H, s | — |
| 4a | — | 115.4 |
| 5 | — | 192.3 |
| 5a | — | 114.4 |
| 6 | — | 162.4 |
| 6-OH | 12.46, 1H, s | — |
| 6a | — | 130.9 |
| 7 | 5.18, 1H, dd, 4.3, 3.1 | 71.3 |
| 8A | 2.51, 1H, dd, 15.0, 4.3 | 33.9 |
| 8B | 2.32, 1H, dd, 15.0, 3.1 | — |
| 9 | — | 72.1 |
| 9-OH | 4.58, 1H, s | — |
| 10 | 4.02, 1H, s | 56.9 |
| 10a | — | 142.4 |
| 11 | 7.40, 1H, s | 120.8 |
| 11a | — | 133.1 |
| 12 | — | 180.7 |
| 12a | — | 132.6 |
| 13A | 1.73, 1H, dq, 14.2, 7.4 | 32.0 |
| 13B | 1.51, 1H, dq, 14.2, 7.4 | — |
| 14 | 1.10, 3H, t, 7.4 | 6.7 |
| 15 | — | 171.1 |
| 16 | 3.69, 3H, s | 52.5 |
| 1' | 5.41, 1H, d, 3.5 | 101.7 |
| 2'a | 1.75, 1H, ddd, 12.8, 11.2, 3.4 | 37.7 |
| 2'e | 2.19, 1H, dd, 12.8, 5.3 | — |
| 3' | 3.71, 1H, ddd, 12.0, 9.0, 5.3 | 69.0 |
| 4' | 3.14, 1H, dd, 9.1, 9.0 | 78.1 |
| 5' | 3.88, 1H, dq, 9.1, 6.2 | 68.8 |
| 6' | 1.36, 3H, d, 6.2 | 17.6 |

Table 3. ^1H and ^{13}C assignments of 9-*epi*-auramycinone (Formula II).

| Site | ^1H | ^{13}C |
|------|-------------------------|-----------------|
| 1 | 7.76, 1H, dd, 7.5, 1.2 | 119.8 |
| 2 | 7.67, 1H, dd, 8.3, 7, 5 | 137.4 |
| 3 | 7.28, 1H, dd, 8.3, 1.2 | 124.8 |
| 4 | — | 162.5 |
| 4-OH | 11.86, 1H, s | — |
| 4a | — | 115.6 |
| 5 | — | 192.7 |
| 5a | — | 114.6 |
| 6 | — | 160.9 |
| 6-OH | 12.76, 1H, s | — |
| 6a | — | 134.1 |
| 7 | 5.40, 1H, t, 7.0 | 64.0 |
| 8A | 2.66, 1H, dd, 13.9, 7.0 | 40.9 |
| 8B | 1.89, 1H, dd, 13.9, 7.1 | — |
| 9 | — | 70.5 |
| 9-OH | 3.49, 1H, brs | — |
| 10 | 3.93, 1H, d, 0.8 | 56.0 |
| 10a | — | 142.1 |
| 11 | 7.51, 1H, d, 0.8 | 120.1 |
| 11a | — | 133.3 |
| 12 | — | 180.9 |
| 12a | — | 132.1 |
| 13 | 1.44, 3H, s | 28.7 |
| 14 | — | 173.0 |
| 15 | 3.90, 3H, s | 52.6 |

References

- Aoki, M., Shirai, H., Nakayama, N., Itezono, Y., Mori, Mitsuko, Satoh, T., Ohishima, S., Watanabe, J., and Yokose, K. 1991. Structural studies on avidinorubicin, a novel anthracycline with platelet aggregation inhibitory activity. *J. Antibiot.* **44**: 635-645.
- Bibb, M.J., Janssen, G.R., and Ward, J.M. 1985. Cloning and analysis of the promoter region of the erythromycin resistance gene (*ErmE*) of *Streptomyces erythraeus*. *Gene* **38**: 215-226.
- Bhuyan, B.K., and Dietz, A. 1965. Fermentation, taxonomic and biological studies of nogalamycin. *Antimicrob. Ag. Chemother.* 1965: 836-844.
- Bhuyan, B.K., and Smith C.G. 1975. In: Antineoplastic and immunosuppressive agents, part II, pp. 623-632. Sartorell, A.C., and Johns, D.G. (eds), Springer-Verlag, Berlin.
- Fujii, I., and Ebizuka, Y. 1997. Anthracycline biosynthesis in *Streptomyces galilaeus*. *Chem. Rev.* **97**: 7: 2511-2523.
- Kawai, H., Hayakawa, Y., Nakagawa, M., Furihata, K., Shimazu, A., Seto, H. and Otake, N. 1987. Arugomycin, a new anthracycline antibiotic-I. Taxonomy, fermentation, isolation and physico-chemical properties. *J. Antibiot.* **19**: 1266-1271.
- Kunnari, T., Tuikkanen, J., Hautala, A., Hakala, J., Ylihonko, K., and Mäntsälä, P. 1997. Isolation and characterization of 8-demethoxy steffimycins and generation of 2,8-demethoxy steffimycins in *Streptomyces steffisburgensis* by the nogalamycin biosynthesis genes. *J. Antibiot.* **50**: 496-501.
- Lomovskaya, N., Doi-Katayama, Y., Filippini, S., Nastro, C., Fonstein, L., Gallo, M., Colombo, A.L. and Hutchinson, C.R. 1998. The *Streptomyces peucetius dpsY* and *dnrX* genes govern early and late steps of daunorubicin and doxorubicin biosynthesis. *J. Bacteriol.* **180**: 9: 2379-2386.
- Madduri, K., Kennedy, J., Rivola, G., Inventi-Solari, A., Filippini, S., Zanuso, G., Colombo, A.L., Gewain, K.M., Occi, J.L., MacNeil, D.J., and Hutchinson, C.R. 1998. Production of antitumor drug epirubicin (4'-epidoxorubicin) and its precursor by a genetically engineered strain of *S. peucetius*. *Nature Biotech.* **16**: 69-74.
- Niemi, J., Ylihonko, K., Hakala, J., Köpio, A., Pärssinen, R., and Mäntsälä, P. 1994. Hybrid anthracycline antibiotics: production of new anthracyclines by cloned genes from *Streptomyces purpurascens* in *Streptomyces galilaeus*. *Microbiol.* **140**: 1351-1358.

Otten, S.L., Liu, X., Ferguson, J., and Hutchinson C.R. 1995. Cloning and characterization of the *Streptomyces peucetius* *dnrQS* genes encoding a daunosamine biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis. *J. Bacteriol.* **177**: 6688–6692.

Rajgarhia, V.B., and Strohl, W.R. 1997. Minimal *Streptomyces* sp. strain C5 daunorubicin polyketide biosynthesis genes required for aklanononic acid biosynthesis. *J. Bacteriol.* **179**: 8: 2690–2696.

~~Strohl, W.R., Bartel, P.L., Li, Y., Connors, N.C., and Woodman, R.H. 1991. Expression of polyketide biosynthesis and regulatory genes in heterologous streptomycetes. *J. Ind. Microbiol.* **7**: 3: 163–174.~~

Torkkell, S., Ylihonko, K., Hakala, J., Skurnik, M., and Mäntsälä, P. 1997. Characterization of *Streptomyces nogalater* genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis. *Mol. Gen. Genet.* **256**: 203–209.

Ward, J.M., Janssen, G.R., Kieser, T., Bibb, M.J., Buttner, M.J., and Bibb, M.J. 1986. Construction and characterization of a series of multicopy promoter–probe plasmid vectors for *Streptomyces* using the aminoglycoside phosphotransferase from Tn5 as indicator. *Mol. Gen. Genet.* **203**: 468–478.

Wiley, P.F., Kelly, R.B., Caron, E.L., Wiley, V.H., Johnson, J.H., MacKellar, F.A. and Mizesak, S.A. 1977. Structure of nogalamycin. *J. Am. Chem. Soc.* **99**: 542–549.

Wiley, P.F., Elrod, D.W., and Marshall, V.P. 1978. Biosynthesis of the anthracycline antibiotics nogalamycin and steffimycin. *B. J. Org. Chem.* **43**: 3457–3461.

Wright, F., and Bibb, M.J. 1992. Codon usage in the G+C –rich *Streptomyces* genome. *Gene* **113**: 55–65.

Ylihonko, K., Hakala, J., Niemi, J., Lundell, J., and Mäntsälä, P. 1994. Isolation and characterization of aclacinomycin A–non–producing *Streptomyces galilaeus* (ATCC 31615) mutants. *Microbiol.* **140**: 1359–1365.

Ylihonko, K., Hakala, J., and Mäntsälä, P. 1995. Process for producing anthracyclines and intermediates thereof. WO 96/10581.

Ylihonko, K., Hakala, J., Kunnari, T., and Mäntsälä, P. 1996a. Production of hybrid anthracycline antibiotics by heterologous expression of *Streptomyces nogalater* nogalamycin biosynthesis genes. *Microbiol.* **142**: 1965–1972.

Ylihonko, K., Tuikkanen, J., Jussila, S., Cong, L., and Mäntsälä, P. 1996b. A gene cluster involved in nogalamycin biosynthesis from *Streptomyces nogalater*: sequence analysis and complementation of early–block mutations in the anthracycline pathway. *Mol. Gen. Genet.* **251**: 113–120.

Ylihonko, K., Hakala, J., and Kunnari, T. 1998. Hybrid anthracyclines from genetically engineered *Streptomyces galilaeus* strains. No. 981062.

Yoshida, M., Fujioka, A., Nakano, K., Yuasa, C., Toko, T., Takeda, S., and Unemi, N. 1996. Activity of menogaril against various malignant lymphoma cell lines and a human lymphoma xenograft in mice. *Anticancer Res.* **16**: 2875-2879.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

- (A) NAME: Galilaeus Oy
- (B) STREET: Kairiskulmantie 10
- (C) CITY: Piispanristi
- (E) COUNTRY: Finland
- (F) POSTAL CODE (ZIP): FIN-20760

(ii) TITLE OF INVENTION: Gene cluster involved in nogalamycin biosynthesis; and its use in production of hybrid antibiotics

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 16020 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (B) STRAIN: *Streptomyces nogalater* ATCC 27451

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (1..1027)
- (D) OTHER INFORMATION: /function= "aminotransferase"
/gene= "snogI"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1192..2073
- (D) OTHER INFORMATION: /function= "dTDP-glucose synthase"
/gene= "snogJ"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (2106..2822)
- (D) OTHER INFORMATION: /function= "aminomethyl transferase"
/gene= "snogA"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (2826..3800)
- (D) OTHER INFORMATION: /function= "polyketide cyclase"
/gene= "snoaM"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3799..5025
- (D) OTHER INFORMATION: /function= "unknown"
/gene= "snogN"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 5088..6356
- (D) OTHER INFORMATION: /function= "hydroxylase"
/gene= "snoaG"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (6334..7209)
- (D) OTHER INFORMATION: /function= "dTDP-4-dehydrorhamnose
reductase"
/gene= "snogC"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (7245..8297)
- (D) OTHER INFORMATION: /function= "dTDP-glucose-4,6-dehydratase"
/gene= "snogK"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 8537..8941
- (D) OTHER INFORMATION: /function= "NAME cyclase"
/gene= "snoaL"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 8992..9699
- (D) OTHER INFORMATION: /function= "unknown"
/gene= "snoK"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (9745..10917)
- (D) OTHER INFORMATION: /function= "glycosyl transferase"
/gene= "snogD"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 11057..11884
- (D) OTHER INFORMATION: /function= "unknown"
/gene= "snoW"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 11928..13200
- (D) OTHER INFORMATION: /function= "glycosyl transferase"
/gene= "snogE"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: complement (13335..13754)
- (D) OTHER INFORMATION: /function= "unknown"
/gene= "snoL"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 13974..14441
- (D) OTHER INFORMATION: /function= "homologous to mtmX of mithramycin
cluster"
/gene= "snoO"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 14532..15377
- (D) OTHER INFORMATION: /function= "C-7 ketoreductase"
/gene= "snoaF"

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: 15450..16020
 (D) OTHER INFORMATION: /function= "unknown"
 /gene= "snoN"

(ix) FEATURE:

(A) NAME/KEY: misc feature
 (B) LOCATION: 3799..3800
 (D) OTHER INFORMATION: /note= "overlapping sequence in the
 genes *snoaM* and *snoG*"

(ix) FEATURE:

(A) NAME/KEY: misc feature
 (B) LOCATION: 6334..6356
 (D) OTHER INFORMATION: /note= "overlapping sequence in the
 genes *snoaG* and *snoG*"

(ix) FEATURE:

(A) NAME/KEY: misc feature
 (B) LOCATION: 13201..13300
 (D) OTHER INFORMATION: /note= "unknown region"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

| | | | | | | |
|------------|------------|------------|------------|------------|-------------|------|
| AGATCTCGTC | CGCCAGTGCC | TCGGTGACCG | GCAACGAGCC | CTTGGCGTAG | CCGAGATGGG | 60 |
| AGAAACCGGT | CATGGTGTGC | ACGGGCCAGG | GATAACTGAT | GTTGAGGGCG | ATGTCGTAGG | 120 |
| AGGCGCGCAG | GGCCTCCAGC | ACCGCGTCCC | GTGCGGGATG | GCGCACCACG | TACACGTAGT | 180 |
| AGACGTGCTC | GTTGCCCTGC | GCGGTCTCTG | GCAGCAGCAG | CCCCGTGTCC | GCCAGGCCCT | 240 |
| CCTCATAGCG | GCGTGCCACC | GCCCGGCGGG | CCTCGATGTA | GGACGGCAAC | CGGGACAGCT | 300 |
| TGCGCCGCAG | GATCTCTGCC | TGTACTTCGT | CCAGCCGGCT | GTTGTGCCCC | GGGGTTTCTGA | 360 |
| CGACGTAGTA | GCGGCTCTCC | ATGCCGTAGT | AGCGCAGCCG | CCGCAGCCGG | TCCGCCACCC | 420 |
| GCTCGTCGTC | GGTGAGCACC | GCGCCGCCGT | CCCCGTACGC | GCCCAGCACC | TTGGTCGGGT | 480 |
| AGAAGGAGAA | CGCGGCCGCG | TCACCGGTCT | AGCCGGCGAG | TCGGCCGTGC | CGGCGCGCCC | 540 |
| CGTGCGCCTG | CGCGCAGTCC | TCCAGGATCA | CCAGGTTGTG | CCGGGCGGCC | AGATCGCGCA | 600 |
| GCGGTGCCAT | GTCCACGCAC | TGCCCGTAGA | GGTGGACCGG | CAGCAGACAC | CGGGTGCGTG | 660 |
| GCGTGAGGAC | GGCCTCCACC | TGGGACGTGT | CCATCAGGTA | GTCCTCCTCG | CGCACGTCCA | 720 |
| CGAAGACGGG | CGTGGCACC | GCCGAGTCGA | TCGCGACGAC | CGTGGGCGCG | GCGGTGTTGG | 780 |
| ACACGGTGAC | GACCTCGTCG | CCGGGCCCGA | CACCCAAGGC | CTGTAACCCC | AGCTTGACGG | 840 |
| CGTTGGTCCC | GTTGTGACG | CCGACGGCAT | GTCCGACGCC | CTGGAATGAG | GCGAACTCGG | 900 |
| ACTCGAAGCC | GCGCAGCTC | TCACCGAGGA | CGAGCCGGCC | GGAGCGGAAC | ACCGTCTCCA | 960 |
| CGGCATCGTG | GATGTCCTCG | CGTTCCAGCT | CGTATTCCGG | CAGATAGTCC | CACACGTGTA | 1020 |
| CGGTCATCGA | GCCCCTCCGG | GATTCTCCCT | GCGAAAAGTC | ACCACTCTAC | GACAACGTTT | 1080 |
| ACCACTCGCT | TTTTCTCAA | CGTCCGCTTG | AGACGGCCCC | GCCTGCTGTG | GCCCCGGGAA | 1140 |
| AGGTGCGGTC | GTTATCATCG | ACTCCGTCTT | CTCATTCGGA | GTTGTTCAG | GGTGAAGGGA | 1200 |
| ATCATTCTCG | CCGGGGGTAC | GGGGAGCAGG | CTCCACCCGA | CGACTCTCGC | GGTGTCCAAG | 1260 |

| | | | | | | |
|-------------|-------------|------------|-------------|------------|------------|------|
| CAGCTTCTCC | CCGTCGGGGA | CAAGCCGATG | ATCTACTACC | CGCTCTCCGT | GCTGATGCTG | 1320 |
| GCCGGCGTCA | CGGACATCCT | CATCATCAGC | ACACCGCACG | AACTCCCCCG | AATGCGCCGT | 1380 |
| CTGTTTCGGCG | ACGGCGCACA | GCTCGGACTC | CGCCTGGCCT | ACGCCGAGCA | GGAGAAACCC | 1440 |
| AGGGGTATCG | CCGAGGCGTT | CCTGATCGGT | GCCGACCACG | TGGGAAGCGA | TGCCGTTGCG | 1500 |
| CTGGCGCTGG | GCGACAACAT | ATTCCACGGG | AGTTCTTTTC | AGGGGGTGCT | GCGCAAGGAA | 1560 |
| GCCGAGGAAT | TGGACGGGTG | TGTCCTGTTC | GGTTATCCGG | TCAAGGATCC | CCAGCGTTAT | 1620 |
| GGAGTCGGCG | AGGCGAACGC | GTCCGGGCGG | CTCGTCTCCA | TCGAGGAGAA | ACCGGTACGC | 1680 |
| CCCCGCTCCA | ACCGGGCCAT | CACCGGACTC | TATTTCTACG | ACAACGAGGT | GGTGGACATC | 1740 |
| GCCCGGCGGC | TGCGCCCCTC | CGCCCGCGGC | GAACTCGAAA | TCACCGACAT | CAACCGTACC | 1800 |
| TACATGGAAC | GAGGCCGGGC | CCGGCTCGTG | GACCTGGGCC | GGGGATTTCG | CTGGCTCGAC | 1860 |
| ACCGGCACAC | CCGAGTCACT | CCTGCAGGCC | TCGCAGTACG | TGTCCGCCCT | GGAGGAACGC | 1920 |
| CAGGGCATCA | GGATCGCCTG | CATCGAGGAG | GTGGCCCTCC | GCATGGGCTT | CATCAACGCC | 1980 |
| CAGGCCTGCT | ACGAACTGGG | CGCGCGCCTG | TCCGGCTCCG | GCTACGGGCA | GTACGTGATG | 2040 |
| GCCATCGCGG | AGGAGTGCAC | GGGGCGGGTG | TGAGCGGCCG | TGCCGGGTGG | GCGAACGGCC | 2100 |
| CGGCCTTACC | CGGCCCCGCG | CACCCCGACG | AACAACCCCC | GGCCGGTCAG | CCCGTCGTCC | 2160 |
| AGGAACTCGG | CCGGGCAGCC | CGCGTCCCTG | AACGCGGCGA | GGTACTCCTC | CCTGGTGAAC | 2220 |
| AGGGTGAGCA | GGTGATCTC | CGTGAACCTG | CGTATCCCGG | TGGCCTCGCC | GACCAGGAAC | 2280 |
| CGCACCTCCA | TGCGGGTCCT | GCGGCCCTGC | CTGGTGGAGT | GGGACACCCG | GGCCACGGTC | 2340 |
| CGGCCCTCAC | CGCGTGCCAG | GTCCCCGGCG | ACGTAGCCCT | CCAGGAACCG | CTCGGGGAAC | 2400 |
| CACCAGGGCT | CCACCACGAG | CACGCEGCCC | GGCACCAGGT | GCGCGGECAT | CGTGCGCACC | 2460 |
| GCCGCCGCA | TGTCCGCGAC | GGTCTCCAGA | TACCCGATGG | AGCAGAACAG | GCAGACCACG | 2520 |
| GCGTCGAAAC | GCCCCGCTCAG | GGCGAAGTCG | CGCATGTCCC | CGGGCCGCAC | CGGCACCCCC | 2580 |
| GGCAGCCGCC | GTTCGGCCAG | GGCCCGCATC | TCGTCCGACA | GCTCCAGGCC | CTCCGTGTGC | 2640 |
| GCGAACAGCC | CGCGGAAGGC | CTCCAGATGG | GCGCCGGTGC | CGCAGGCGAC | GTCGAGCAGC | 2700 |
| GAACGCGCCC | CGGGCCGACG | GGACCTGATC | TCCGCGGTGA | CCCGTTCGGC | CTCGTCCGCC | 2760 |
| CAGCTCTTTC | CCCCGGCTGCG | GTAGACCATC | TCGTACACGT | CCGCCAGTTC | CCGGCCGTAC | 2820 |
| ACGCGTCAGT | CCTCGTCCAC | CAGGGCGACC | GCCCGGGTCC | ACCCGGCGCC | GGCGCCGGCG | 2880 |
| ACCTTGACCG | GGAAGCAGGA | GACGGGGAAC | CCGAAGGAGA | CCGGCAGGEG | GTCGAGGTC | 2940 |
| GCCAGCCGCT | CGATCTGGCA | GTAGTCCCGG | TCCCGGCCCA | CCACGTGCGG | AGGGGCAAGC | 3000 |
| ACCGATCGGT | CGCGGGTGGG | CGCGTACCGG | TCCGATGAGT | GGGCGAAGGG | CGGGTCCAGG | 3060 |
| CTGAAGGCAT | CGGTCCCGAT | CACCCGGACC | CCGTGGTCTGA | GAAGCATCCG | TACCGCGGGC | 3120 |
| CCGTGAGAC | CGGCGAAGTC | CGTGAAGTAG | CGCGGGGTGC | CCGCGTGCCG | CTGGGCACCG | 3180 |
| GTGTGCAGCA | GCACGATGTC | CCCCGGCCGC | AACGCGCACC | CGGTCCGGGC | CAGTTCCTTC | 3240 |
| TCCAGGCGCG | CGGCGCTCAC | GGTGCCCGTC | GGAGCGTCGG | TGAGGTCCAG | CACCACCCCG | 3300 |

| | | | | | | |
|------------|------------|------------|-------------|------------|------------|------|
| CGCCCGAAGA | ACCACTCCAG | CGGCATCTGG | TCGATGTGGC | GGGGGACGCC | GTCCCCGTAC | 3360 |
| AGCGCGCGCG | AACCATAGTG | CGACGGCGCG | TCGACGTGCG | TGCCGGTGTG | CGTGGTCAGC | 3420 |
| GTGATCCTGT | CCAGTGACAG | GAAGTCGCCG | TCCGGCAGTT | CGTCCGGAGA | GAAGTCGACA | 3480 |
| CCGAAGTGCT | CGCGCATCTC | CGCGCACATG | TGTTCCGCGC | CCTGCCGGGG | CGTGAGGACG | 3540 |
| TCGTGCACCA | CCGGGTCGGG | CTCGTACTGT | GAGGAATCCA | CCGGTGACGA | AAGGTCGATG | 3600 |
| AGCCGCACGC | GCACCTCCGG | GTTCGTAGAC | GGGCTCGGCT | GACGCAGCGC | GGGTACGACG | 3660 |
| CTGACACGCC | CCTCTTGACG | TGGCCTGGAA | GCTGGTTCTGA | CGGGCGGGCA | CCGCACGCGA | 3720 |
| CGGCCGGCGC | CGCACCGGCG | CCGTCCCGGC | CGAGCGGGAA | TCCAGGGAGG | GTATAGCGGC | 3780 |
| GCGCCCCACG | CTGCCGTCAT | GGTGATGAAA | CTGACGGACA | GCGAGCTGGG | GCGTGCGCTG | 3840 |
| CTCTCGCTGC | GTGGTTACCA | GTGGCTCCGC | GGCATCCACC | ACGATCCCTA | CGCCCTGCTG | 3900 |
| CTGCGCGCCG | AGAGCGACGA | TCCGGCGCAG | CTCGGCCGGC | TGCTGCGTGA | ACGCGGCCCG | 3960 |
| CTCCACCGCA | GCGACACCGG | CACCTGGGTC | ACCGCGGACC | ATGCGACGGC | CTCCCGGCTG | 4020 |
| CTCGCCGACC | CGCGCTTCGT | GCTGCGCCGC | CCGCCGGCCG | GGCCCCCCAC | CGGCACCGGG | 4080 |
| GACGTCATGC | CGTGGAAGA | GGCCACGCTG | AGCGACCTGC | TGCCCCCTGA | CGAGGCGCGC | 4140 |
| CTGACGACCG | ACCGGGCACG | GTGCCGCCGG | CTCGGCGCGA | CCGCCGCGCG | GATCGCGGCG | 4200 |
| GACGGTCCCG | TCGCGACGCG | ACTCGCGGAC | CTGGCCGGGG | CCCAGCCGA | ACAGGTGCGC | 4260 |
| TCAACGGGCC | ACTTCGACCT | CAGGGCCGAC | TACGCCCTCC | CGTACGCGGT | CGAGCCGGCC | 4320 |
| TGCGCGCTGC | TCGGCCTGCC | GGCCGGGCAG | TGTTCCCTCT | TCGGCGCCTT | CTCCCCGGCC | 4380 |
| GTCTGCTCG | ACGCGACGGT | CGTACCGCCC | CGCCTTCCGG | AGGCGCGCGC | CCTGATCGCC | 4440 |
| TCCACGGCGG | AACTGACCGC | CCTCTGGCCG | CGGCTGGCCC | CGAGCCTGTC | GAAGACCGTC | 4500 |
| CCGGAGGACG | AAGCGCCGGA | CCTCTTCCTG | CTGACGGCCG | TGTTACTCGT | ACCGGCCGTC | 4560 |
| GTCCACCTGG | TCTGCGAGGC | GGTCGCCGCC | CTGTGCGACG | ACCCCGGGCA | GGCCGGGCTG | 4620 |
| CTCAGGGACG | ACCCGGTACT | CGCCGCACCG | GCGGTGAGG | AGACGCTGCG | CCACGCACCG | 4680 |
| CCCGCCCGTC | TGTTACCCCT | CCACGCGACC | GGACCGGAGC | GCGTCGCGGA | CGTCGACCTC | 4740 |
| CCCGCGGGCG | CCGAGGTCGC | CGTCGTCTGT | GCGGCGGCGC | ACCGCGATCC | CTCCTGGTGC | 4800 |
| CCGGACCCCG | ACCGCTTCGA | CCTCACCAGG | AACGAGCGGC | ATCTGGCACT | GCCGCCGGAT | 4860 |
| CTGCCGCTGG | GGGCGCTCGC | CCCGCTGCTG | CGCGTCTGCG | CGACCGCGGC | CGTCGCGGCC | 4920 |
| CTCGCGGCCG | GACTCCTCCC | GCTGCGGGCC | GTCGGCCCCG | CCGTACGACG | GCTGCGTGCC | 4980 |
| CCGGTCACCC | GGTCCGTGCT | GCGCTTCCCC | GTCGCCCCGT | GCTGAGCAGC | CCCTCCTCAC | 5040 |
| GTCATCCCCG | GCCCGCCTTC | CCCCGCCCCG | AACGGAAGGG | ACTCTCCATG | GACAACCGCG | 5100 |
| AGACCGTACG | ACCGGTGAGC | GTCTGCCGGG | TCTGCGGCGG | CAACGACTGG | CAGGACGTCG | 5160 |
| TGGACTTCGG | TGACGTTCCC | CTCGCCAACG | GCTTCCTGTC | CCCGGCCGAC | TCCTACGAGA | 5220 |
| ACGAGCGCCG | CTACCCGCTG | GGCGTCCTGT | CCTGCCGCGC | CTGCCGGCTG | ATGAGCCTGA | 5280 |
| CCCACGTGGT | CGACCCCGAG | GTGCTGTACC | GCGACTACGC | CTACACCACC | CCCGACTCCG | 5340 |

| | | | | | | |
|-------------|-------------|------------|------------|-------------|------------|------|
| AAATGATCAC | CCAGCACATG | CGGCACATCA | CCGCGCTGTG | CCGCACCCGT | TTCGAGCTTC | 5400 |
| CCCCGGACAG | CCTCGTCGTG | GAGCTGGGCA | GCAATACCGG | CCGTCAGCTC | ATGGCCTTCC | 5460 |
| GCGAAGCGGG | GATGCGCACC | CTGGGCGTGG | ACCCCGCGCG | CAACCTCACG | GACGTCGCCC | 5520 |
| GGCGCAACGG | CATCGAGACC | TTCCCCGACT | TCTTCTCCCA | CGACGTGGCC | CGCACCATCC | 5580 |
| GGCGCGACCA | CGGGCAGGCG | CGGCTCGTGC | TGGGACGGCA | TGTCTTCGCC | CACATCGACG | 5640 |
| ACGTGTCGGA | CATCGCGGCC | GGCGTACGCG | AACTCCTGTC | TCCCGACGGG | GTGTTCGCGA | 5700 |
| TCGAGGTGCC | GTACGTTCTG | GACCTGCTGG | AGAAGGTCGC | GTTTCGACACC | ATCTACCACG | 5760 |
| AGCACTTGTC | GTACTTCACC | ATGCGGTCCT | TCGTCACCCT | CTTCGCGCGC | CACGGGCTGC | 5820 |
| GGGTGCTCGA | CGTGGAGCGG | TTCGGCGTGC | ACGGCGGATC | GGTCCTCGTC | TTCGTGGGCC | 5880 |
| ACGAGGACGG | CCCCTGGCCC | GAACGTCCCT | CCGTCCCCGA | ACTGCTGCGC | GTGGAACGGC | 5940 |
| AGCGGGGCCT | CTACGACGAC | GCCACCTACC | GCACGTTCCG | GCAGCGGATC | GAGCGGGTGC | 6000 |
| GCACCGAACT | GCCGGAAGTG | CTGCGCTCCC | TCGTGGCCCA | GGGCAAGCGC | ATCGTCGGCT | 6060 |
| ACGGTGCTCC | GGCCAAGGGC | AACACCATCC | TCACGGTGTG | CGGGCTCGGC | CTGAAGGAGC | 6120 |
| TGGAATACTG | CACCGACACC | ACCGAGCTGA | AGCAGGGCAG | GGTGCTGCCC | GGCACCACAC | 6180 |
| TACCGGTGCA | CGCTCCCGAG | CACGCCAAGG | AACACATCCC | CGACTACTAC | CTGTTGCTCG | 6240 |
| CCTGGAACCTA | CGCCACGGAG | ATCCTCGACA | AGGAGACGGC | CTTCCGGGAC | AACGGCGGCC | 6300 |
| GGTTCATCGT | GCCCATCCCC | CGCCGTCGA | TCCTCACGTC | CCCGTCAGGT | TCCTGAGGCG | 6360 |
| CCCGCCGGGC | AGCAGCTGAC | GCATCGCCTC | GCGCAGGGCT | GCACGCCAGT | CGCGGGGCGG | 6420 |
| TGCGACGCCG | ACCAGCCGCC | AGCGGTCTGT | CCCGAGCACC | GTGCACGCCG | GCCGGGGCGC | 6480 |
| CGGGCCCCGG | CGGTCCGGCC | TCGCCACCGG | GCGCACCCGT | TCCGGGTCCG | CGCCCGCCAG | 6540 |
| CCGGAACACC | TCCCGGGCCA | GCTCGTACCA | GGTGGCCGCC | CCGGCGTTGG | TGGCGTGGAA | 6600 |
| GATCCCGCGC | GCCC GGCTGT | GCGGCGTGCG | GGCCAGCGTC | ACCAGCAGCC | GGGCCACGTC | 6660 |
| ACCGGCCCCAC | GTCGGCTGCC | CCCACTGGTC | GTTGACGACG | TCGACATGGC | CGTCGTCCGG | 6720 |
| GGCACGCTCC | AGCATCGTGC | GCACGAAGCT | GCGGCCCTGC | CCGCCGTAGA | GCCACGCCGT | 6780 |
| GCGCAACACG | GTGCCCCTAT | CCGGCAGCAG | CGACAGCACG | GCCC GTTCCC | CGGCCAGTTT | 6840 |
| GCTGCGGCCG | TACACCGTGC | GCGGGCCCCG | AGCGTCCGAC | TCGCCGTAAG | GGCTGCGGGT | 6900 |
| GTCGCCCCGG | AAGACGTAGT | CGGTGAGAC | GTGGATCAGC | CGTACGCCGT | GGCGCGCACA | 6960 |
| GCGGCGGGCC | AGCAGCCGGG | GCCCGCCGCC | GTTGACGCGC | ATCGCCTCCG | CCCACCGCGA | 7020 |
| CTCGGCGCCG | TCCACGTCCG | TGAAGGCGGC | GCAGTTGACC | ACCACCGGCG | GCCGGTGCGC | 7080 |
| GGCGAACCGG | GCGTCCACCG | CCCGCCCGTC | GGTGATGTCC | AGCGCGCGCC | GCCCCAGTAC | 7140 |
| CACCGCCTCG | GCGGCGGGCC | GGCTCCTGCC | GGTCTCCGCC | AGGGCCGCGG | TCAGGTGCCG | 7200 |
| GGCGAGCATG | CCTTCTCCTC | CGGTGACCAG | CACGCGCATC | CCGCTCACCG | GACCCCGGGG | 7260 |
| ACGACGGTGG | ACGTACCGCC | CGGCGCCGTG | ACTCCCCGCT | TGAGCGGCTC | CCACCAGGAC | 7320 |
| CGGTTCTCGC | GGTACCACTG | GACCGTCGAG | CGCAGCCCCG | AGGAGAACTC | CCGCGCCGGA | 7380 |

| | | | | | | |
|------------|------------|------------|------------|------------|------------|------|
| CGGTAGCCCA | GTTCTCAG | GGCCCTGCCC | CAGTCCAGGC | TGTAACGCAG | GTCGTGCCCC | 7440 |
| TTGCGGTCGG | GCACGTGCCG | GACGCTGCTC | CAGTCCGCCC | CGCACAGCTC | CAGCAACATA | 7500 |
| CCCACCAGCT | CCCGGTTGGA | GAGCTCCCGG | CCGCCGCCGA | TGTGGTACAC | ACCGCCGGGC | 7560 |
| CGGCCCGCGG | TGCGCACCAG | GTCCACGCCC | CGGCAGTGGT | CCTCCACGTG | CAGCCACTCC | 7620 |
| CGCACGTTCC | GGCCGTCCCC | GTACAGCGGC | ACCGGCAGCC | CGTCCAACAA | GTTGGTGACG | 7680 |
| AAGCGCGGGA | TGAGCTTCTC | CGGGTGCTGA | CGCGGGCCGT | AGTTGTTGGA | ACAGCGGGTC | 7740 |
| ACCCGCACGT | CCAGGCCGTG | CGTGCGGTGG | CAGGCGAACG | CCATCAGGTC | GGCCGACGCC | 7800 |
| TTGGAGGCGG | CGTACGGGGA | GTTGGGGCTC | AGCGGGTGCT | CCTCCGGCCA | GGAACCGGAC | 7860 |
| GCGATGGAGC | CGTAGACCTC | GTCCGTGGAC | ACCAGGACGA | AGGGCTCCAC | GCCGTGGCGC | 7920 |
| AGCGCGGCGT | CCAGCAGCCG | CTGGGTGCCG | ACGACGTTGG | TCAGCACGAA | GTCGTGCGCC | 7980 |
| GCGCGGATGG | ACCGGTCGAC | GTGCGACTCC | GCGGCGAAGT | GGACGACCTG | GTCGCTGTGT | 8040 |
| GCCATCAGCT | CGTCGACCAG | CTCGGCGTCG | AGGATGTCGC | CCCGCACGAA | GCGCAGCCGG | 8100 |
| TCACCGCGTA | CCGCGTCCAG | GTTGCTGAGG | TTGCCCGCGT | ACGTCAGTTT | GTCGAGGACG | 8160 |
| GTGACGCGTA | CCGCCGGGGC | CCCCGCTCCG | GGGGCCCGGT | TCTCCAGCAG | CATGCGCACA | 8220 |
| TAGGCCGAGC | CGATGAAACC | GACCGCGCCG | GTGACCAGGA | TGTTACGTC | CGTCGTCGCG | 8280 |
| GAGGTGTGCG | ACGCCATGGG | TTCCCTCCAT | CCGTCGGGTG | CCGTGGGGCG | GAGTGCGCCC | 8340 |
| CCTCGACCCA | GCGTCGGGGG | CGGCCGTGGA | GGAGCGGTTG | AGCTTCGGCG | CAGCGGCGGC | 8400 |
| TCGACCGGCG | GCGGCCGGCG | TCGCCGGACT | CCAACGGTTC | TCGACGGAAC | GACCAACGGC | 8460 |
| CCTGGCGAGA | CTGCCCGGAC | AGCCCGGCCG | AGAGAGGGAG | GACCCGTTGA | GCCGTCAGAC | 8520 |
| AGAGATCGTC | CGCCGGATGG | TGAGCGCCTT | CAACACCGGC | AGGACCGACG | ACGTGGACGA | 8580 |
| GTACATCCAC | CCCGACTACC | TCAATCCGGC | CACCTTGGA | CACGGCATCC | ACACCGGGCC | 8640 |
| CAAGGCGTTC | GCCCAGCTGG | TCGGCTGGGT | GCGGGCGACG | TTCTCCGAGG | AAGCCCGCCT | 8700 |
| GGAGGAGGTG | CGGATCGAGG | AGCGCGGCCC | GTGGGTCAAG | GCCTACCTCG | TGCTCTACGG | 8760 |
| CCGCCACGTC | GGCCGGCTTG | TCGGTATGCC | GCCCACCGAC | CGGCGCTTCT | CCGGTGAACA | 8820 |
| GGTGACCTG | ATGCGCATCG | TCGACGGGAA | GATCCGCGAC | CACCGGGACT | GGCCCGACTT | 8880 |
| CCAGGGGACG | CTGCGCCAGC | TCGGCGACCC | GTGGCCCGAC | GACGAGGGCT | GGCGTCCGTG | 8940 |
| ACCGTCCCTG | AAACCGCACC | CGACGAGACA | TCAGACCAGG | AAGGATGGCT | CATGCCGGAT | 9000 |
| CCCGGCGGCC | CGACCACGGC | CGAGAACCTG | TCGAAGGAGG | CTGTCCGCTT | CTACCGCGAG | 9060 |
| CAGGGTTACG | TGCACATCCC | GCGCGTCTTG | TCGGAGACGG | AGGTGACCGC | CTTCCGGGCC | 9120 |
| GCCTGTGAGG | AGGTCCTGGA | GAAGGAGGGC | CGCGAGATCT | CCGGCATCGC | CCTGCGGCTG | 9180 |
| GCCGGCGCGC | CCCTGCGGGT | CTACAGCAGC | GACATCTTGG | TCAAGGAGCC | CAAGCGCACC | 9240 |
| CTGCCCACCC | TGGTCCACGA | CGACGAGACG | GGAATGCCGC | TGAACGAGCT | GAGTGCCACG | 9300 |
| CTGACGGCCT | GGATCGCGCT | GACGGACGTA | CCCGTCGAAC | GCGGCTGCAT | GAGCTACGTG | 9360 |
| CCGGGCTCCC | ATCTCAGGGC | CCGCGAGGAC | CGGCAGGAGC | ACATGACCAG | CTTCGCCGAG | 9420 |

| | |
|---|-------|
| TTCCGGGACC TCGCGGACGT GTGGCCCGAT TACCCGTGGC AGCCGCGCGT CGCCGTGCCC | 9480 |
| GTCCGCGCCG GAGACGTCGT GTTCCACCAT TGCCGTACCG TCCACATGGC CGAAGCCAAC | 9540 |
| ACCAGCGACT CGGTCCGCAT GGCAGATGGC GTCGTCTACA TGGACGCGGA CGCCACCTAC | 9600 |
| CGGCCGGGCG TCCAGGACGG CCACCTGTCC CGCCTGTGCG CGGGAGATCC ACTCGAAGGC | 9660 |
| GAGCTGTTCC CCCTGGTCAC GGCAGGCACA CGGCAGTGAG GTCCGCCGTT CCCGGEGGTC | 9720 |
| GCGGGACCGC CGGGGACGGC ACCGTCAGCC GGCCAGCGCC ACGAGCTTGG CGGEGGTCTC | 9780 |
| GGCCGGCGGC GGCATCTCGC TCATCTCCTG CCGCACCCGC AGGGCCGCCT CCCGCAACCC | 9840 |
| CGCGTCGTCC AGCAGCCGTC GGCAGTGTCT GGCACCCAGC GATCCCGCCT CGGCATCGAA | 9900 |
| CCCGATGCCC AGCCCGGTCA GCACATCGCG GTTGGTGTCC TGGTAGGAGC CGTGCGGGAT | 9960 |
| GACGCACTGC GGGACGCCCG CGGCCAGGGC CGTCAGCAGT GTGCCGCTGC CCCCCTGATG | 10020 |
| GATGATCGCG TCGCACGTCT CCAGCAGCGC GCCCAGCGGA ATCCACTCCA CCACCGGTAC | 10080 |
| GTTCGCGGGC AGTTCACCGA GCAGGGCCAG GTCGCCGCCG CCCAGGGTCA GCACGAATC | 10140 |
| CGCGTCCACG TCCGCCACTT CGGAGAACAG CGGGGCCAGC TTGGCGATGC CGCCCGACAG | 10200 |
| CGCGTCGATG GAGCCCAGCG TCACCGCGAT ACGCCGCCGG CCGGCCGCGG GCGGCAGCCA | 10260 |
| GTCCGGCAGC ACCGCTCCGC CGTTGTAGGG GACGTACCGC ATCGGCCAGG CACCCGGGGA | 10320 |
| GCGCCGGTCC TCCGGGAGCA GCGCTCCAC GCTCGGGGGT GTCGTCTGTA GCCGCACGGA | 10380 |
| ACCGGTCGGC TCGCCGGTGA GCGCGTGGCG CTCGTAGTCC TTGGACATCG CCCGCCGGAT | 10440 |
| GAGCGCGCCG AGCCCGGGT CGCTGTCCGC GGGAGCCAGC GGCAGCTCTA CGCAGGGCAG | 10500 |
| TTGCAGCGCT GCCGCCGTC GCGGGCCCGC GCGCTGTGTC GGAGTGTGCA CGACGAGGTC | 10560 |
| GGGCCGCCAG CTCGGGGGGT TCCGCAGGGC CCCGTGAGCG GCGACGGCGC ATACCCGGGC | 10620 |
| GAACATCTCG GCGAAGAAGC CCTCGCCCAG CCCCTCGGAG TGCATCGGGT CGGTGACGTC | 10680 |
| GGTGTCTGCG GGCACGAACA GCTTCGCGTA GTTCACGCCG GGCAGACAGT CCACGGCGCA | 10740 |
| CAGCCCGGCC TCCCGACGG CGCGGATGTC GCGCCCGGTG GCGTAGCGGA CCTCGTGGCC | 10800 |
| GAGAGCGCGC AGCGCCTGTG CCAGCGGCAC CGTCGGCAGG ATGTGGCTGA GCCCGGGTGA | 10860 |
| AGTGATGAAC AACGCACGCA TGATGCCCC TGTTGACAT GAACCTGGAA CACGCATCCT | 10920 |
| GACGGCGCCT TCTGTTGCTC CGGTCGACGC CCGGTCGACA GGCCCTCGTA CAGCCCGCCG | 10980 |
| GGGGCCGGTC CGGCCACGAC GCAGGCTCCA GCGGACGTCG ACGGCGGGGA CGCAGCGTGG | 11040 |
| TCGCCGGGAG GCATCGATGA CAGTATGGT AACCGGAGCG ACAGGAAAGC TCGGCGGGCA | 11100 |
| CGTCGTCACC GGGCTACTGG CCGCGGGGCG CCGGGTGGCG GCGCTGAGCG GCACAGCGCA | 11160 |
| CCGGTCCGGC CTGCGCGGGC GCGCGGAGAT CAGAGGGGCG GAGCTGAGCG GCGCGGAGAC | 11220 |
| CTACGAGCGG ATGCTGGACG GTGTGGAAGC CGTCTACCTG TTCCCCGTCC CGGAGACCGC | 11280 |
| CGCGGCGTTC GCCGGGGCCG CGCGACGGGC CGGTGTCCGG CGGATCGTGG TGCTCTCCTC | 11340 |
| GGACTCCGTC ACCGACGGCA CCGACACCGG AGGACACCGG CGCGTGGAAC TGGCCGTGGA | 11400 |
| GGACACGGGG CTCGAGTGGA CCCATGTGCG CCGCGGCGAG TTCGCGCTCA ACAAGGTCAC | 11460 |

| | | | | | | |
|------------|-------------|-------------|------------|------------|------------|-------|
| CCTGTGGGCG | CCGTCGATCC | GCGCGGAGGG | CGTCGTCCGG | TCCGCGTATC | CGGACGCCCC | 11520 |
| GGTGGCCCCG | GTGCACGAGG | CCGACGTCGC | GGCCGTCCGC | GTGACCGCGC | TGCTGAAGGA | 11580 |
| GGGGCACGCC | GGCCGCGCCT | ACAGCGTGAC | CGGACCGCAG | GCCCTCACCC | AGCGCGAACA | 11640 |
| GGTCCGCGCG | GTAGGGGAGG | GGCTCGGCCG | GTCCCTCGCC | TTCGTGAGG | TGACCCCCGG | 11700 |
| GCAGGCGCGG | GCCGACCTGA | CCGCCCAGGG | GCTGCCCGCG | CCCATCGCCG | ACTACGTCCT | 11760 |
| CGCCTTCCAA | GCCGGGTGGA | CCGAGCGGCC | CGCCCCCGCC | CGGCCGACCG | TGCGGGAGGT | 11820 |
| CACCGGCCGG | CCCGCCCGCA | CGCTCGCCCA | GTGGGCCGCC | GACCACCGAG | CGGACTTCCG | 11880 |
| GTGACCGGAG | ACCGCGTCCA | CCGCGCCACG | ACAGAAAGGC | GACGCCCGTG | CGCGTACTGC | 11940 |
| TGACGTCCTT | CGCCATGGAC | GCCCACTTCT | GCACCGCCGT | GCCGCTGGCG | TGGGCACTGC | 12000 |
| GGTCGGCCGG | GCACGAGGTA | CGGGTGGCCG | GCCAGCCCGC | GCTCACCTCC | ACCATCACGG | 12060 |
| GAGCCGGCCT | GACCGCCGTG | CCGGTCGGCC | GCGACCACAC | GCACGGCAGC | CTCCTGGGCC | 12120 |
| GGGTCGGCAG | CGACATCCTC | GCCCTGCACG | ACGAGGCGGA | CTACCTGGAG | GCCCGTCACG | 12180 |
| ACGCCCTGGG | CTTCGAGTTC | CTCAAAGGGC | ACAACACGGT | GATGTCCGCG | TTGTTCTACT | 12240 |
| CGCAGATCAA | CAACGACTCG | ATGGTCGACG | ACCTGGTGGA | CTTCGCCCGT | CACTGGCGGC | 12300 |
| CCGACCTGGT | CGTCTGGGAG | CCGTTACCT | TCGCGGGCGC | CGTGGCCGCG | CGGGCCTCGG | 12360 |
| GCGCCGCCCA | CGCCCGCCTG | CTGTCTTCC | CCGACCTGTT | CCTCAGCACG | CGCCGCCTCT | 12420 |
| TCCTGGAGCG | CATGGCGCGC | CAGGAGCCCG | AGCATCACGA | CGACACACTC | GCCGAATGGC | 12480 |
| TCGACTGGAC | CCTTGGCCGG | CACGGCCACT | CCTTCGACGA | GGAGATCGTC | ACGGGGCAGT | 12540 |
| GGTCCATCGA | CCAGACCCCC | GCCCCCGTGC | GGCTCGACGC | CGGCGGTCCC | ACCGTGCCGA | 12600 |
| TGCGGTACGT | CCCCCTACAGC | GGACTGGTGC | CCACAGTGGT | GCCCGACTGG | CTGCGCAGGC | 12660 |
| CGCCCGAGCG | GCCACGGGTC | CTGGTCACCC | TCGGCATCAC | CTCACGGCGG | GTGAAGTCCT | 12720 |
| TCCTCGCCGT | CTCCGTGGAC | GACCTTTTCG | AGGCCGTGGC | CGGGCTCGGC | GTGAGGTGG | 12780 |
| TCGCCACCCT | CGACGCCGAC | CAGCGGGAGC | TGCTGGGGCG | CGTGCCGGAC | CACTTCCGCA | 12840 |
| TCGTGAGCA | CGTGCCGCTG | GACGCCGTTC | TGCCGACCTG | CTCGGCGATC | GTCCACCACG | 12900 |
| GCGGAGCCGG | CACCTGGTCG | ACGGCCGCCG | TGTACGGGGT | GCCGCAGGTC | TCCCTGGGCT | 12960 |
| CGATGTGGGA | CCACTTCTAC | CGGGCCCCGTC | GCCTGGAGGA | ACTCGGGGCG | GGGCTGCGGC | 13020 |
| TGCCCTCCGG | CGAGCTGACT | GCCGAGGGGC | TGCGCACCCG | GCTGGAGAGG | GTGCTCGGCG | 13080 |
| AGCCCTCCTT | CGGCACCGCC | GCGCAGGCGC | TGAGCGACAC | CATCGCGGCG | GAACCCAGCC | 13140 |
| CCAGCGAGGT | CGTGCCGCTC | CTGGAGGAGC | TGACCGGACG | GCACCGTCCC | GGCACCCGGG | 13200 |
| NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | 13260 |
| NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | NNNNNNNNNN | CCGTCCGGGC | CCCTCGCCGG | 13320 |
| TGAGGGAGCC | CGGATCACAG | TCCGTCCGGC | ACCACGCCCA | GGTCCCGGAA | CAGCGGGGAG | 13380 |
| AAGTTGAAGA | CGTCCCAGTG | CTCCACGACC | TTGCCGGCTT | CGGAGAAGCG | CAGCTCCTCC | 13440 |
| AAGTAGGTCC | AGCGGACCTT | GCGGCCGGTG | GGGGCGATGC | CCATGAACAC | GCCCTGGTGC | 13500 |

| | | | | | | |
|------------|------------|------------|-------------|------------|-------------|-------|
| GTGGCCGAGC | AGGTGATCCG | CAGCATCACG | CGGTCGCCCT | CGCCCACGAT | GCTCCGCACG | 13560 |
| TCCAGACGAA | GGTCCGGGAA | GGCCTCCACC | GCGCTGTTCA | TACGCCGTAC | GACCTCCTCG | 13620 |
| GCGCTCACCG | GTTTGTCTC | GTCGTCTAG | TGGACGACGT | CGGGTGCCCA | GTGCGCGACC | 13680 |
| ACCCCGGAGA | CGTCCCACCG | GTTCCATGCG | GCCACCATCT | CCAGGCAGCG | TTCCTTGTTT | 13740 |
| GCGGTCTGTT | ACATGTCGAC | TCCTTGAAGG | CCCGGGACTA | CTGCTCACGC | GCCAGCCTTC | 13800 |
| CAACCCGCCC | CGGAAAAGCG | GTGCACGACC | GCTGGAGCCC | GCACCGGAAC | CTGCGCGGCG | 13860 |
| GAGCTGAACG | GGGTTTCGAG | CCGTTACCA | AGGACCTGCC | GCAGCCTGTT | ACGGCACACC | 13920 |
| CTGACGCCTC | GCTCCGCGCG | GGACGCGCCC | GCCGGGAGGA | AGGACACACC | ACCATGTCGG | 13980 |
| TACGCACCGA | TCAGACGGCG | GCACCGGAAG | ACCGAGCGGC | GGCCACGGAT | CCCGGGTTTCG | 14040 |
| GGCACCTGTA | CGCGCAGGTG | CAGCAGTTCT | ACGCCCCGCA | GATGCAGCTC | CTCGACTCCG | 14100 |
| GCGCGGCCGA | GGAGTGGGCC | GCCACCTTCA | CCGAGGACGG | CACGTTGCGC | CGGCCCTCCT | 14160 |
| CGCCGGAACC | GGCACGCGGC | CACGCCGAAC | TGGCCGCCGG | CGCCCGCGCC | GCCGCCGAAC | 14220 |
| GCCTCGCCGC | CGAGGGCCTT | TCGCACCGGC | ACGTCATCGG | CATGACCGCG | GTACGCCGGG | 14280 |
| AACCCGACGG | CAGCGTGTTT | GTACGCAGCT | ACGCCCAGGT | CTTCGCCACC | CGCCGCGGGG | 14340 |
| AAGCTCCCCG | GCTGCATCTG | ATCTGCGTCT | GCGAGGACGT | GCTCGTGCGG | GAGGGGCCGG | 14400 |
| GGCTGAAGGT | GCGGGAACGG | GTTGTACGCG | ACGACGCGTG | AGGGCGGTCT | ACCCGCCGGC | 14460 |
| CGAGCCGCAC | CTCTGCCACC | CCCTCGGCAC | GCCAGCGGGC | GTCGAGTCCG | CTGCGAGAGG | 14520 |
| GCGCACTTAG | CGTGCGAGCC | ATGACTGACT | CGACAGGTCC | CCGCCCGGTG | CCCGCCATGT | 14580 |
| CACCCGCCCC | CAGCCCCACG | CCTTCCCCCG | GCCCCGCCCC | CGGGAGCGAA | CCCGCGCCGC | 14640 |
| TCGCCGTGAT | CGTCACCGGC | GGCGGTTCCG | GTATCGGCGG | GGCCACCGCC | CGCGCCTTCG | 14700 |
| CCGCTCAGGG | TGCGAAGGTG | CTCGTCGTCG | GCCGTACCGA | GGACGCGCTC | GCGCAGACCG | 14760 |
| CCGAGGGCTG | TGCGGACATG | CGTGTGCTCG | TCGCCGACGT | GGCCTCGCCC | GACGGGCCGC | 14820 |
| AGGCGGTCGT | CAACGCCGCC | CTGCGGGAGT | TCGGGAGGAT | CGACGTCTTG | GTCAACAACG | 14880 |
| CTGCCGTGGC | GGGCATGGAG | ACCCTGCAGA | CCGTCGACCG | GGACGCCGTG | GCACGGCAGT | 14940 |
| TCGGCACCAA | TCTGACGGCT | CCCCTCTTCC | TCGTCCAGTC | CGCACTCGGC | GCGCTGGAGA | 15000 |
| AGTCGCGCGG | CATCGTCGTC | AACGTGGGGA | CCGCCGCGAC | CCTGGGCCTG | CGCGCCGCCC | 15060 |
| CGACCGGCGC | GCTGTACGGG | GCGAGCAAGG | TGGCCCTCGA | CTACCTGACC | CGGACCTGGG | 15120 |
| CCGTCGAACT | GGCCCCCGG | GGCATCCGTG | TCGTGCGCGT | GGCACCCGGG | GTGATCGACA | 15180 |
| CGGGCATCGG | CGTCCGCATG | GGCATGACCC | CGGAGGGGCTA | CCGGGAGTTC | CTGACCGGCA | 15240 |
| TGGGCGGCAG | GGTGCCCGTG | GGGCGGGTGG | GCCGTCCGGA | GGAGGTGGCG | TGGTGGATCG | 15300 |
| TCCAGCTCGC | CCGCCCGGAG | GCCGGCTACG | CGACGGGCAT | GGTCGTCCCC | GTCGACGGCG | 15360 |
| GGCTGTGCTG | GGTCTGACCG | GACAAGGAAG | GAAATACCGC | AGGAAGGAAG | TACCGCAGCA | 15420 |
| AGGAAATACC | GCAGGAAGGA | GATATCGCCG | TGCAGGAAAC | CGAACCCGGC | GTCCCCGCGG | 15480 |
| ACCTGCCCGC | CGAGAGCGAC | CCTGCCGCCC | TGGAGCGCCT | CGCCGCACGG | TACCGGCGGG | 15540 |

```

ACGGCTACGT CCACGTCCCC GCGTCCTCG ACGCCGGGGA GGTCGCCGAA TACCTGGCCG 15600
AGGCCCCGTCG GCTCCTCGCC CACGAGGAGT CCGTGCGCTG GGGCTCCGGC GCCGGCACCG 15660
TCATGGACTA CGTCGCCGAC GCCCAGCTCG GCAGCGACAC GATGCGCCGC CTTGCCACCC 15720
ACCCGCGCAT CGCCGCCCTC GCCGAGTACC TGGCCGGCTC GCCCCTGAGG CTGTTCAAGC 15780
TGGAGGTGCT GCTCAAGGAG AACAAGGAGA AGGACGCCTC GGTCGCCACC GCCCCGCACC 15840
ACGATGCGTT CGCCTTCCCC TTCTCCACCG CCGGCACCGC CCTGACGGCG TGGGTGCGCG 15900
TGGTCGACGT CCCGGTGGAA CGCGGCTGCA TGACCTTCGT CCCC GGATCA CACCTGCTGC 15960
CGGATCCCCA TACCGGCGAC GAGCCGTGGG CCGGGGCCTT CACCCGGCCG GGAGAGATCT 16020

```

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 342 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snogI"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

Met Thr Val His Val Trp Asp Tyr Leu Pro Glu Tyr Glu Leu Glu Arg
1           5           10           15

Glu Asp Ile His Asp Ala Val Glu Thr Val Phe Arg Ser Gly Arg Leu
20           25           30

Val Leu Gly Glu Ser Val Arg Gly Phe Glu Ser Glu Phe Ala Ser Phe
35           40           45

Gln Gly Val Gly His Ala Val Gly Val Asp Asn Gly Thr Asn Ala Val
50           55           60

Lys Leu Gly Leu Gln Ala Leu Gly Val Gly Pro Gly Asp Glu Val Val
65           70           75           80

Thr Val Ser Asn Thr Ala Ala Pro Thr Val Val Ala Ile Asp Ser Ala
85           90           95

Gly Ala Thr Pro Val Phe Val Asp Val Arg Glu Glu Asp Tyr Leu Met
100          105          110

Asp Thr Ser Gln Val Glu Ala Val Leu Thr Pro Arg Thr Arg Cys Leu
115          120          125

Leu Pro Val His Leu Tyr Gly Gln Cys Val Asp Met Ala Pro Leu Arg
130          135          140

Asp Leu Ala Ala Arg His Asn Leu Val Ile Leu Glu Asp Cys Ala Gln
145          150          155          160

Ala His Gly Ala Arg Arg His Gly Arg Leu Ala Gly Ser Thr Gly Asp
165          170          175

Ala Ala Ala Phe Ser Phe Tyr Pro Thr Lys Val Leu Gly Ala Tyr Gly
180          185          190

```

Asp Gly Gly Ala Val Leu Thr Asp Asp Glu Arg Val Ala Asp Arg Leu
 195 200 205
 Arg Arg Leu Arg Tyr Tyr Gly Met Glu Ser Arg Tyr Tyr Val Val Glu
 210 215 220
 Thr Pro Gly His Asn Ser Arg Leu Asp Glu Val Gln Ala Glu Ile Leu
 225 230 235 240
 Arg Arg Lys Leu Ser Arg Leu Pro Ser Tyr Ile Glu Ala Arg Arg Ala
 245 250 255
 Val Ala Arg Arg Tyr Glu Glu Gly Leu Ala Asp Thr Gly Leu Leu Leu
 260 265 270
 Pro Arg Thr Ala Gln Gly Asn Glu His Val Tyr Tyr Val Tyr Val Val
 275 280 285
 Arg His Pro Arg Arg Asp Ala Val Leu Glu Ala Leu Arg Ala Ser Tyr
 290 295 300
 Asp Ile Ala Leu Asn Ile Ser Tyr Pro Trp Pro Val His Thr Met Thr
 305 310 315 320
 Gly Phe Ser His Leu Gly Tyr Ala Lys Gly Ser Leu Pro Val Thr Glu
 325 330 335
 Ala Leu Ala Asp Glu Ile
 340

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 298 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (D) OTHER INFORMATION: /note= "translate of snogJ"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Val Lys Gly Ile Ile Leu Ala Gly Gly Thr Gly Ser Arg Leu His Pro
 1 5 10 15
 Thr Thr Leu Ala Val Ser Lys Gln Leu Leu Pro Val Gly Asp Lys Pro
 20 25 30
 Met Ile Tyr Tyr Pro Leu Ser Val Leu Met Leu Ala Gly Val Thr Asp
 35 40 45
 Ile Leu Ile Ile Ser Thr Pro His Glu Leu Pro Arg Met Arg Arg Leu
 50 55 60
 Phe Gly Asp Gly Ala Gln Leu Gly Leu Arg Leu Ala Tyr Ala Glu Gln
 65 70 75 80
 Glu Lys Pro Arg Gly Ile Ala Glu Ala Phe Leu Ile Gly Ala Asp His
 85 90 95
 Val Gly Ser Asp Ala Val Ala Leu Ala Leu Gly Asp Asn Ile Phe His
 100 105 110
 Gly Ser Ser Phe Gln Gly Val Leu Arg Lys Glu Ala Glu Glu Leu Asp
 115 120 125

Gly Cys Val Leu Phe Gly Tyr Pro Val Lys Asp Pro Gln Arg Tyr Gly
 130 135 140
 Val Gly Glu Ala Asn Ala Ser Gly Arg Leu Val Ser Ile Glu Glu Lys
 145 150 155 160
 Pro Val Arg Pro Arg Ser Asn Arg Ala Ile Thr Gly Leu Tyr Phe Tyr
 165 170 175
 Asp Asn Glu Val Val Asp Ile Ala Arg Arg Leu Arg Pro Ser Ala Arg
 180 185 190

 Gly Glu Leu Glu Ile Thr Asp Ile Asn Arg Thr Tyr Met Glu Arg Gly
 195 200 205

 Arg Ala Arg Leu Val Asp Leu Gly Arg Gly Phe Ala Trp Leu Asp Thr
 210 215 220

 Gly Thr Pro Glu Ser Leu Leu Gln Ala Ser Gln Tyr Val Ser Ala Leu
 225 230 235 240

 Glu Glu Arg Gln Gly Ile Arg Ile Ala Cys Ile Glu Glu Val Ala Leu
 245 250 255

 Arg Met Gly Phe Ile Asn Ala Gln Ala Cys Tyr Glu Leu Gly Ala Arg
 260 265 270

 Leu Ser Gly Ser Gly Tyr Gly Gln Tyr Val Met Ala Ile Ala Glu Glu
 275 280 285

 Cys Thr Gly Arg Val
 290

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 238 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snogA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Val Tyr Gly Arg Glu Leu Ala Asp Val Tyr Glu Met Val Tyr Arg Ser
 1 5 10 15
 Arg Gly Lys Ser Trp Ala Asp Glu Ala Glu Arg Val Thr Ala Glu Ile
 20 25 30
 Arg Ser Arg Arg Pro Gly Ala Arg Ser Leu Leu Asp Val Ala Cys Gly
 35 40 45
 Thr Gly Ala His Leu Glu Ala Phe Arg Gly Leu Phe Ala His Thr Glu
 50 55 60
 Gly Leu Glu Leu Ser Asp Glu Met Arg Ala Leu Ala Glu Arg Arg Leu
 65 70 75 80
 Pro Gly Val Pro Val Arg Pro Gly Asp Met Arg Asp Phe Ala Leu Ser
 85 90 95
 Gly Arg Phe Asp Ala Val Val Cys Leu Phe Cys Ser Ile Gly Tyr Leu
 100 105 110

Glu Thr Val Ala Asp Met Arg Ala Ala Val Arg Thr Met Ala Ala His
 115 120 125
 Leu Val Pro Gly Gly Val Leu Val Val Glu Pro Trp Trp Phe Pro Glu
 130 135 140
 Arg Phe Leu Glu Gly Tyr Val Ala Gly Asp Leu Ala Arg Gly Glu Gly
 145 150 155 160
 Arg Thr Val Ala Arg Val Ser His Ser Thr Arg Gln Gly Arg Arg Thr
 165 170 175
 Arg Met Glu Val Arg Phe Leu Val Gly Glu Ala Thr Gly Ile Arg Glu
 180 185 190
 Phe Thr Glu Ile Asp Leu Leu Thr Leu Phe Thr Arg Glu Glu Tyr Leu
 195 200 205
 Ala Ala Phe Glu Asp Ala Gly Cys Pro Ala Glu Phe Leu Asp Asp Gly
 210 215 220
 Leu Thr Gly Arg Gly Leu Phe Val Gly Val Arg Gly Ala Gly
 225 230 235

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 324 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note="translate of snoaM"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Thr Ala Ala Trp Gly Ala Pro Leu Tyr Pro Pro Trp Ile Pro Ala
 1 5 10 15
 Arg Pro Gly Arg Arg Arg Cys Gly Ala Gly Arg Arg Val Arg Cys Pro
 20 25 30
 Pro Val Glu Pro Ala Ser Arg Pro Arg Gln Glu Gly Arg Val Ser Val
 35 40 45
 Val Pro Ala Leu Arg Gln Pro Ser Pro Ser Thr Asn Pro Glu Val Arg
 50 55 60
 Val Arg Leu Ile Asp Leu Ser Ser Pro Val Asp Ser Ser Gln Tyr Glu
 65 70 75 80
 Pro Asp Pro Val Val His Asp Val Leu Thr Pro Arg Gln Gly Ala Glu
 85 90 95
 His Met Cys Ala Glu Met Arg Glu His Phe Gly Val Glu Phe Ser Pro
 100 105 110
 Asp Glu Leu Pro Asp Gly Glu Phe Leu Ser Leu Asp Arg Ile Thr Leu
 115 120 125
 Thr Thr His Thr Gly Thr His Val Asp Ala Pro Ser His Tyr Gly Ser
 130 135 140
 Arg Ala Leu Tyr Gly Asp Gly Val Pro Arg His Ile Asp Gln Met Pro
 145 150 155 160

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Leu | Glu | Trp | Phe | Phe | Gly | Arg | Gly | Val | Val | Leu | Asp | Leu | Thr | Asp | Ala |
| | | | | 165 | | | | | 170 | | | | | 175 | |
| Pro | Thr | Gly | Thr | Val | Ser | Ala | Ala | Arg | Leu | Glu | Lys | Glu | Leu | Ala | Arg |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Thr | Gly | Cys | Ala | Leu | Arg | Pro | Gly | Asp | Ile | Val | Leu | Leu | His | Thr | Gly |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Ala | Gln | Arg | His | Ala | Gly | Thr | Pro | Arg | Tyr | Phe | Thr | Asp | Phe | Ala | Gly |
| | 210 | | | | | 215 | | | | | 220 | | | | |
| Leu | Asp | Gly | Pro | Ala | Val | Arg | Met | Leu | Leu | Asp | His | Gly | Val | Arg | Val |
| 225 | | | | | 230 | | | | | 235 | | | | 240 | |
| Ile | Gly | Thr | Asp | Ala | Phe | Ser | Leu | Asp | Ala | Pro | Phe | Gly | His | Ile | Ile |
| | | | | 245 | | | | | 250 | | | | | 255 | |
| Asp | Arg | Tyr | Arg | Ala | Thr | Gly | Asp | Arg | Ser | Val | Leu | Trp | Pro | Ala | His |
| | | | 260 | | | | | 265 | | | | | 270 | | |
| Val | Val | Gly | Arg | Glu | Arg | Glu | Tyr | Cys | Gln | Ile | Glu | Arg | Leu | Ala | Asn |
| | | 275 | | | | | 280 | | | | | 285 | | | |
| Leu | Asp | Arg | Leu | Pro | Val | Ser | Phe | Gly | Phe | Arg | Val | Cys | Cys | Phe | Pro |
| | 290 | | | | | 295 | | | | | 300 | | | | |
| Val | Lys | Val | Ala | Gly | Ala | Gly | Ala | Gly | Trp | Thr | Arg | Ala | Val | Ala | Leu |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 |
| Val | Asp | Glu | Asp | | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 408 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snogN"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Val | Met | Lys | Leu | Thr | Asp | Ser | Glu | Leu | Gly | Arg | Ala | Leu | Leu | Ser |
| 1 | | | | 5 | | | | | 10 | | | | | 15 | |
| Leu | Arg | Gly | Tyr | Gln | Trp | Leu | Arg | Gly | Ile | His | His | Asp | Pro | Tyr | Ala |
| | | | 20 | | | | | 25 | | | | | 30 | | |
| Leu | Leu | Leu | Arg | Ala | Glu | Ser | Asp | Asp | Pro | Ala | Gln | Leu | Gly | Arg | Leu |
| | | 35 | | | | | 40 | | | | | 45 | | | |
| Leu | Arg | Glu | Arg | Gly | Arg | Leu | His | Arg | Ser | Asp | Thr | Gly | Thr | Trp | Val |
| | 50 | | | | | 55 | | | | | 60 | | | | |
| Thr | Ala | Asp | His | Ala | Thr | Ala | Ser | Arg | Leu | Leu | Ala | Asp | Pro | Arg | Phe |
| 65 | | | | | 70 | | | | | 75 | | | | | 80 |
| Val | Leu | Arg | Arg | Pro | Pro | Ala | Gly | Pro | Ala | Thr | Gly | Thr | Gly | Asp | Val |
| | | | 85 | | | | | 90 | | | | | | 95 | |
| Met | Pro | Trp | Glu | Glu | Ala | Thr | Leu | Ser | Asp | Leu | Leu | Pro | Leu | Asp | Glu |
| | | | 100 | | | | | 105 | | | | | 110 | | |

| | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ala | Arg | Leu | Thr | Thr | Asp | Arg | Ala | Arg | Cys | Arg | Arg | Leu | Gly | Ala | Thr |
| | | 115 | | | | | 120 | | | | | 125 | | | |
| Ala | Ala | Arg | Ile | Ala | Ala | Asp | Gly | Pro | Val | Ala | Thr | Arg | Leu | Ala | Asp |
| | 130 | | | | | 135 | | | | | 140 | | | | |
| Leu | Ala | Gly | Ala | Arg | Ala | Glu | Gln | Val | Arg | Ser | Thr | Gly | His | Phe | Asp |
| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
| Leu | Arg | Ala | Asp | Tyr | Ala | Leu | Pro | Tyr | Ala | Val | Glu | Pro | Ala | Cys | Ala |
| | | | 165 | | | | | 170 | | | | | | 175 | |
| Leu | Leu | Gly | Leu | Pro | Ala | Gly | Gln | Cys | Ser | Leu | Phe | Gly | Ala | Phe | Ser |
| | | | 180 | | | | | 185 | | | | | 190 | | |
| Pro | Ala | Val | Leu | Leu | Asp | Ala | Thr | Val | Val | Pro | Pro | Arg | Leu | Pro | Glu |
| | | 195 | | | | | 200 | | | | | 205 | | | |
| Ala | Arg | Ala | Leu | Ile | Ala | Ser | Thr | Ala | Glu | Leu | Thr | Ala | Leu | Trp | Pro |
| | 210 | | | | | 215 | | | | | 220 | | | | |
| Arg | Leu | Ala | Pro | Ser | Leu | Ser | Lys | Thr | Val | Pro | Glu | Asp | Glu | Ala | Pro |
| 225 | | | | | 230 | | | | | 235 | | | | | 240 |
| Asp | Leu | Phe | Leu | Leu | Thr | Ala | Val | Leu | Leu | Val | Pro | Ala | Val | Val | His |
| | | | 245 | | | | | 250 | | | | | | 255 | |
| Leu | Val | Cys | Glu | Ala | Val | Ala | Ala | Leu | Ser | His | Asp | Pro | Gly | Gln | Ala |
| | | 260 | | | | | | 265 | | | | | 270 | | |
| Gly | Leu | Leu | Arg | Asp | Asp | Pro | Val | Leu | Ala | Ala | Pro | Ala | Val | Glu | Glu |
| | 275 | | | | | 280 | | | | | | 285 | | | |
| Thr | Leu | Arg | His | Ala | Pro | Pro | Ala | Arg | Leu | Phe | Thr | Leu | His | Ala | Thr |
| | 290 | | | | | 295 | | | | | 300 | | | | |
| Gly | Pro | Glu | Arg | Val | Ala | Asp | Val | Asp | Leu | Pro | Ala | Gly | Ala | Glu | Val |
| 305 | | | | | 310 | | | | | 315 | | | | | 320 |
| Ala | Val | Val | Val | Ala | Ala | Ala | His | Arg | Asp | Pro | Ser | Trp | Cys | Pro | Asp |
| | | | | 325 | | | | | 330 | | | | | 335 | |
| Pro | Asp | Arg | Phe | Asp | Leu | Thr | Arg | Asn | Glu | Arg | His | Leu | Ala | Leu | Pro |
| | | | 340 | | | | | 345 | | | | | 350 | | |
| Pro | Asp | Leu | Pro | Leu | Gly | Ala | Leu | Ala | Pro | Leu | Leu | Arg | Val | Cys | Ala |
| | | 355 | | | | | 360 | | | | | 365 | | | |
| Thr | Ala | Ala | Val | Ala | Ala | Leu | Ala | Ala | Gly | Leu | Leu | Pro | Leu | Arg | Ala |
| | 370 | | | | | 375 | | | | | 380 | | | | |
| Val | Gly | Pro | Pro | Val | Arg | Arg | Leu | Arg | Ala | Pro | Val | Thr | Arg | Ser | Val |
| 385 | | | | | 390 | | | | | 395 | | | | | 400 |
| Leu | Arg | Phe | Pro | Val | Ala | Pro | Cys | | | | | | | | |
| | | | | 405 | | | | | | | | | | | |

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 422 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (D) OTHER INFORMATION: /note= "translate of snoag"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Met Asp Asn Arg Glu Thr Val Arg Pro Val Ser Val Cys Arg Val Cys
1 5 10 15

Gly Gly Asn Asp Trp Gln Asp Val Val Asp Phe Gly Asp Val Pro Leu
20 25 30

Ala Asn Gly Phe Leu Ser Pro Ala Asp Ser Tyr Glu Asn Glu Arg Arg
35 40 45

Tyr Pro Leu Gly Val Leu Ser Cys Arg Ala Cys Arg Leu Met Ser Leu
50 55 60

Thr His Val Val Asp Pro Glu Val Leu Tyr Arg Asp Tyr Ala Tyr Thr
65 70 75 80

Thr Pro Asp Ser Glu Met Ile Thr Gln His Met Arg His Ile Thr Ala
85 90 95

Leu Cys Arg Thr Arg Phe Glu Leu Pro Pro Asp Ser Leu Val Val Glu
100 105 110

Leu Gly Ser Asn Thr Gly Arg Gln Leu Met Ala Phe Arg Glu Ala Gly
115 120 125

Met Arg Thr Leu Gly Val Asp Pro Ala Arg Asn Leu Thr Asp Val Ala
130 135 140

Arg Arg Asn Gly Ile Glu Thr Phe Pro Asp Phe Phe Ser His Asp Val
145 150 155 160

Ala Arg Thr Ile Arg Arg Asp His Gly Gln Ala Arg Leu Val Leu Gly
165 170 175

Arg His Val Phe Ala His Ile Asp Asp Val Ser Asp Ile Ala Ala Gly
180 185 190

Val Arg Glu Leu Leu Ser Pro Asp Gly Val Phe Ala Ile Glu Val Pro
195 200 205

Tyr Val Leu Asp Leu Leu Glu Lys Val Ala Phe Asp Thr Ile Tyr His
210 215 220

Glu His Leu Ser Tyr Phe Thr Met Arg Ser Phe Val Thr Leu Phe Ala
225 230 235 240

Arg His Gly Leu Arg Val Leu Asp Val Glu Arg Phe Gly Val His Gly
245 250 255

Gly Ser Val Leu Val Phe Val Gly His Glu Asp Gly Pro Trp Pro Glu
260 265 270

Arg Pro Ser Val Pro Glu Leu Leu Arg Val Glu Arg Gln Arg Gly Leu
275 280 285

Tyr Asp Asp Ala Thr Tyr Arg Thr Phe Ala Gln Arg Ile Glu Arg Val
290 295 300

Arg Thr Glu Leu Pro Glu Leu Leu Arg Ser Leu Val Ala Gln Gly Lys
305 310 315 320

Arg Ile Val Gly Tyr Gly Ala Pro Ala Lys Gly Asn Thr Ile Leu Thr
325 330 335

Val Cys Gly Leu Gly Leu Lys Glu Leu Glu Tyr Cys Thr Asp Thr Thr
340 345 350

Glu Leu Lys Gln Gly Arg Val Leu Pro Gly Thr His Ile Pro Val His
 355 360 365
 Ala Pro Glu His Ala Lys Glu His Ile Pro Asp Tyr Tyr Leu Leu Leu
 370 375 380
 Ala Trp Asn Tyr Ala Thr Glu Ile Leu Asp Lys Glu Thr Ala Phe Arg
 385 390 395 400
 Asp Asn Gly Gly Arg Phe Ile Val Pro Ile Pro Arg Pro Ser Ile Leu
 405 410 415
 Thr Ser Pro Ser Gly Ser
 420

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 291 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (D) OTHER INFORMATION: /note= "translate of snogC"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Leu Ala Arg His Leu Thr Ala Ala Leu Ala Glu Thr Gly Arg Ser
 1 5 10 15
 Arg Pro Ala Ala Glu Ala Val Val Leu Gly Arg Arg Ala Leu Asp Ile
 20 25 30
 Thr Asp Gly Arg Ala Val Asp Ala Ala Phe Ala Ala His Arg Pro Arg
 35 40 45
 Val Val Val Asn Cys Ala Ala Phe Thr Asp Val Asp Gly Ala Glu Ser
 50 55 60
 Arg Trp Ala Glu Ala Met Arg Val Asn Gly Gly Gly Pro Arg Leu Leu
 65 70 75 80
 Ala Arg Arg Cys Ala Arg His Gly Val Arg Leu Ile His Val Ser Thr
 85 90 95
 Asp Tyr Val Phe Pro Gly Asp Thr Arg Ser Pro Tyr Gly Glu Ser Asp
 100 105 110
 Ala Pro Gly Pro Arg Thr Val Tyr Gly Arg Ser Lys Leu Ala Gly Glu
 115 120 125
 Arg Ala Val Leu Ser Leu Leu Pro Asp Thr Gly Thr Val Val Arg Thr
 130 135 140
 Ala Trp Leu Tyr Gly Gly Gln Gly Arg Ser Phe Val Arg Thr Met Leu
 145 150 155 160
 Glu Arg Ala Pro Asp Asp Gly His Val Asp Val Val Asn Asp Gln Trp
 165 170 175
 Gly Gln Pro Thr Trp Ala Gly Asp Val Ala Arg Leu Leu Val Thr Leu
 180 185 190
 Ala Arg Thr Pro Pro Asp Arg Ala Arg Gly Ile Phe His Ala Thr Asn
 195 200 205

Ala Gly Ala Ala Thr Trp Tyr Glu Leu Ala Arg Glu Val Phe Arg Leu
 210 215 220

Ala Gly Ala Asp Pro Glu Arg Val Arg Pro Val Ala Thr Ala Asp Arg
 225 230 235 240

Pro Gly Pro Ala Pro Arg Pro Ala Cys Thr Val Leu Gly His Asp Arg
 245 250 255

Trp Arg Leu Val Gly Val Ala Pro Pro Arg Asp Trp Arg Ala Ala Leu
 260 265 270

Arg Glu Ala Met Arg Gln Leu Leu Pro Gly Gly Arg Leu Arg Asn Leu
 275 280 285

Thr Gly Thr
 290

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 350 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snogK"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Met Ala Ser His Thr Ser Ala Thr Thr Asp Val Asn Ile Leu Val Thr
 1 5 10 15

Gly Ala Val Gly Phe Ile Gly Ser Ala Tyr Val Arg Met Leu Leu Glu
 20 25 30

Asn Arg Ala Pro Gly Ala Gly Ala Pro Ala Val Arg Val Thr Val Leu
 35 40 45

Asp Lys Leu Thr Tyr Ala Gly Asn Leu Thr Asn Leu Asp Ala Val Arg
 50 55 60

Gly Asp Arg Leu Arg Phe Val Arg Gly Asp Ile Leu Asp Ala Glu Leu
 65 70 75 80

Val Asp Glu Leu Met Ala His Ser Asp Gln Val Val His Phe Ala Ala
 85 90 95

Glu Ser His Val Asp Arg Ser Ile Arg Ala Ala Asp Asp Phe Val Leu
 100 105 110

Thr Asn Val Val Gly Thr Gln Arg Leu Leu Asp Ala Ala Leu Arg His
 115 120 125

Gly Val Glu Pro Phe Val Leu Val Ser Thr Asp Glu Val Tyr Gly Ser
 130 135 140

Ile Ala Ser Gly Ser Trp Pro Glu Glu His Pro Leu Ser Pro Asn Ser
 145 150 155 160

Pro Tyr Ala Ala Ser Lys Ala Ser Ala Asp Leu Met Ala Phe Ala Cys
 165 170 175

His Arg Thr His Gly Leu Asp Val Arg Val Thr Arg Cys Ser Asn Asn
 180 185 190

Tyr Gly Pro Arg Gln His Pro Glu Lys Leu Ile Pro Arg Phe Val Thr
 195 200 205
 Asn Leu Leu Asp Gly Leu Pro Val Pro Leu Tyr Gly Asp Gly Arg Asn
 210 215 220
 Val Arg Glu Trp Leu His Val Glu Asp His Cys Arg Gly Val Asp Leu
 225 230 235 240
 Val Arg Thr Ala Gly Arg Pro Gly Gly Val Tyr His Ile Gly Gly Gly
 245 250 255
 Arg Glu Leu Ser Asn Arg Glu Leu Val Gly Met Leu Leu Glu Leu Cys
 260 265 270
 Gly Ala Asp Trp Ser Ser Val Arg His Val Pro Asp Arg Lys Gly His
 275 280 285
 Asp Leu Arg Tyr Ser Leu Asp Trp Gly Arg Ala Arg Glu Glu Leu Gly
 290 295 300
 Tyr Arg Pro Ala Arg Glu Phe Ser Ser Gly Leu Arg Ser Thr Val Gln
 305 310 315 320
 Trp Tyr Arg Glu Asn Arg Ser Trp Trp Glu Pro Leu Lys Arg Gly Val
 325 330 335
 Thr Ala Pro Gly Gly Thr Ser Thr Val Val Pro Gly Val Arg
 340 345 350

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 134 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (D) OTHER INFORMATION: /note= "translate of *snoaL*"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Val Ser Ala Phe Asn Thr Gly Arg Thr Asp Asp Val Asp Glu Tyr
 1 5 10 15
 Ile His Pro Asp Tyr Leu Asn Pro Ala Thr Leu Glu His Gly Ile His
 20 25 30
 Thr Gly Pro Lys Ala Phe Ala Gln Leu Val Gly Trp Val Arg Ala Thr
 35 40 45
 Phe Ser Glu Glu Ala Arg Leu Glu Glu Val Arg Ile Glu Glu Arg Gly
 50 55 60
 Pro Trp Val Lys Ala Tyr Leu Val Leu Tyr Gly Arg His Val Gly Arg
 65 70 75 80
 Leu Val Gly Met Pro Pro Thr Asp Arg Arg Phe Ser Gly Glu Gln Val
 85 90 95
 His Leu Met Arg Ile Val Asp Gly Lys Ile Arg Asp His Arg Asp Trp
 100 105 110
 Pro Asp Phe Gln Gly Thr Leu Arg Gln Leu Gly Asp Pro Trp Pro Asp
 115 120 125

Asp Glu Gly Trp Arg Pro
130

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 235 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snoK"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Met Pro Asp Pro Gly Gly Pro Thr Thr Ala Glu Asn Leu Ser Lys Glu
1 5 10 15

Ala Val Arg Phe Tyr Arg Glu Gln Gly Tyr Val His Ile Pro Arg Val
20 25 30

Leu Ser Glu Thr Glu Val Thr Ala Phe Arg Ala Ala Cys Glu Glu Val
35 40 45

Leu Glu Lys Glu Gly Arg Glu Ile Ser Gly Ile Ala Leu Arg Leu Ala
50 55 60

Gly Ala Pro Leu Arg Val Tyr Ser Ser Asp Ile Leu Val Lys Glu Pro
65 70 75 80

Lys Arg Thr Leu Pro Thr Leu Val His Asp Asp Glu Thr Gly Leu Pro
85 90 95

Leu Asn Glu Leu Ser Ala Thr Leu Thr Ala Trp Ile Ala Leu Thr Asp
100 105 110

Val Pro Val Glu Arg Gly Cys Met Ser Tyr Val Pro Gly Ser His Leu
115 120 125

Arg Ala Arg Glu Asp Arg Gln Glu His Met Thr Ser Phe Ala Glu Phe
130 135 140

Arg Asp Leu Ala Asp Val Trp Pro Asp Tyr Pro Trp Gln Pro Arg Val
145 150 155 160

Ala Val Pro Val Arg Ala Gly Asp Val Val Phe His His Cys Arg Thr
165 170 175

Val His Met Ala Glu Ala Asn Thr Ser Asp Ser Val Arg Met Ala His
180 185 190

Gly Val Val Tyr Met Asp Ala Asp Ala Thr Tyr Arg Pro Gly Val Gln
195 200 205

Asp Gly His Leu Ser Arg Leu Ser Pro Gly Asp Pro Leu Glu Gly Glu
210 215 220

Leu Phe Pro Leu Val Thr Ala Gly Thr Arg Gln
225 230 235

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 390 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

- (D) OTHER INFORMATION: /note= "translate of snogD"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Met Arg Val Pro Gly Ser Cys Arg Thr Gly Gly Ile Met Arg Ala Leu
1 5 10 15

Phe Ile Thr Ser Pro Gly Leu Ser His Ile Leu Pro Thr Val Pro Leu
20 25 30

Ala Gln Ala Leu Arg Ala Leu Gly His Glu Val Arg Tyr Ala Thr Gly
35 40 45

Gly Asp Ile Arg Ala Val Ala Glu Ala Gly Leu Cys Ala Val Asp Val
50 55 60

Ser Pro Gly Val Asn Tyr Ala Lys Leu Phe Val Pro Asp Asp Thr Asp
65 70 75 80

Val Thr Asp Pro Met His Ser Glu Gly Leu Gly Glu Gly Phe Phe Ala
85 90 95

Glu Met Phe Ala Arg Val Ser Ala Val Ala Val Asp Gly Ala Leu Arg
100 105 110

Thr Ala Arg Ser Trp Arg Pro Asp Leu Val Val His Thr Pro Thr Gln
115 120 125

Gly Ala Gly Pro Leu Thr Ala Ala Ala Leu Gln Leu Pro Cys Val Glu
130 135 140

Leu Pro Leu Gly Pro Ala Asp Ser Glu Pro Gly Leu Gly Ala Leu Ile
145 150 155 160

Arg Arg Ala Met Ser Lys Asp Tyr Glu Arg His Gly Val Thr Gly Glu
165 170 175

Pro Thr Gly Ser Val Arg Leu Thr Thr Thr Pro Pro Ser Val Glu Ala
180 185 190

Leu Leu Pro Glu Asp Arg Arg Ser Pro Gly Ala Trp Pro Met Arg Tyr
195 200 205

Val Pro Tyr Asn Gly Gly Ala Val Leu Pro Asp Trp Leu Pro Pro Ala
210 215 220

Ala Gly Arg Arg Arg Ile Ala Val Thr Leu Gly Ser Ile Asp Ala Leu
225 230 235 240

Ser Gly Gly Ile Ala Lys Leu Ala Pro Leu Phe Ser Glu Val Ala Asp
245 250 255

Val Asp Ala Glu Phe Val Leu Thr Leu Gly Gly Gly Asp Leu Ala Leu
260 265 270

Leu Gly Glu Leu Pro Ala Asn Val Pro Val Val Glu Trp Ile Pro Leu
275 280 285

Gly Ala Leu Leu Glu Thr Cys Asp Ala Ile Ile His His Gly Gly Ser
290 295 300

Gly Thr Leu Leu Thr Ala Leu Ala Ala Gly Val Pro Gln Cys Val Ile
305 310 315 320

Pro His Gly Ser Tyr Gln Asp Thr Asn Arg Asp Val Leu Thr Gly Leu
325 330 335

Gly Ile Gly Phe Asp Ala Glu Ala Gly Ser Leu Gly Ala Glu Gln Cys
340 345 350

Arg Arg Leu Leu Asp Asp Ala Gly Leu Arg Glu Ala Ala Leu Arg Val
355 360 365

Arg Gln Glu Met Ser Glu Met Pro Pro Pro Ala Glu Thr Ala Ala Lys
370 375 380

Leu Val Ala Leu Ala Gly
385 390

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 275 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snow"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Met Thr Val Leu Val Thr Gly Ala Thr Gly Asn Val Gly Arg His Val
1 5 10 15

Val Thr Gly Leu Leu Ala Ala Gly Arg Arg Val Arg Ala Leu Thr Arg
20 25 30

Thr Pro Asp Arg Ser Gly Leu Pro Gly Gly Ala Glu Ile Thr Gly Gly
35 40 45

Asp Leu Thr Arg Pro Glu Thr Tyr Glu Arg Met Leu Asp Gly Val Glu
50 55 60

Ala Val Tyr Leu Phe Pro Val Pro Glu Thr Ala Ala Phe Ala Gly
65 70 75 80

Ala Ala Arg Arg Ala Gly Val Arg Arg Ile Val Val Leu Ser Ser Asp
85 90 95

Ser Val Thr Asp Gly Thr Asp Thr Gly Gly His Arg Arg Val Glu Leu
100 105 110

Ala Val Glu Asp Thr Gly Leu Glu Trp Thr His Val Arg Pro Gly Glu
115 120 125

Phe Ala Leu Asn Lys Val Thr Leu Trp Ala Pro Ser Ile Arg Ala Glu
130 135 140

Gly Val Val Arg Ser Ala Tyr Pro Asp Ala Arg Val Ala Pro Val His
145 150 155 160

Glu Ala Asp Val Ala Ala Val Ala Val Thr Ala Leu Leu Lys Glu Gly
165 170 175

His Ala Gly Arg Ala Tyr Ser Val Thr Gly Pro Gln Ala Leu Thr Gln
 180 185 190
 Arg Glu Gln Val Arg Ala Val Gly Glu Gly Leu Gly Arg Ser Leu Ala
 195 200 205
 Phe Val Glu Val Thr Pro Gly Gln Ala Arg Ala Asp Leu Thr Ala Gln
 210 215 220
 Gly Leu Pro Ala Pro Ile Ala Asp Tyr Val Leu Ala Phe Gln Ala Gly
 225 230 235 240
 Trp Thr Glu Arg Pro Ala Pro Ala Arg Pro Thr Val Arg Glu Val Thr
 245 250 255
 Gly Arg Pro Ala Arg Thr Leu Ala Gln Trp Ala Ala Asp His Arg Ala
 260 265 270
 Asp Phe Arg
 275

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: over 424 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note="translate of snogE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Val Arg Val Leu Leu Thr Ser Phe Ala Met Asp Ala His Phe Cys Thr
 1 5 10 15
 Ala Val Pro Leu Ala Trp Ala Leu Arg Ser Ala Gly His Glu Val Arg
 20 25 30
 Val Ala Gly Gln Pro Ala Leu Thr Ser Thr Ile Thr Gly Ala Gly Leu
 35 40 45
 Thr Ala Val Pro Val Gly Arg Asp His Thr His Gly Ser Leu Leu Gly
 50 55 60
 Arg Val Gly Ser Asp Ile Leu Ala Leu His Asp Glu Ala Asp Tyr Leu
 65 70 75 80
 Glu Ala Arg His Asp Ala Leu Gly Phe Glu Phe Leu Lys Gly His Asn
 85 90 95
 Thr Val Met Ser Ala Leu Phe Tyr Ser Gln Ile Asn Asn Asp Ser Met
 100 105 110
 Val Asp Asp Leu Val Asp Phe Ala Arg His Trp Arg Pro Asp Leu Val
 115 120 125
 Val Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala Ala Arg Ala Ser
 130 135 140
 Gly Ala Ala His Ala Arg Leu Leu Ser Phe Pro Asp Leu Phe Leu Ser
 145 150 155 160
 Thr Arg Arg Leu Phe Leu Glu Arg Met Ala Arg Gln Glu Pro Glu His
 165 170 175

His Asp Asp Thr Leu Ala Glu Trp Leu Asp Trp Thr Leu Gly Arg His
180 185 190

Gly His Ser Phe Asp Glu Glu Ile Val Thr Gly Gln Trp Ser Ile Asp
195 200 205

Gln Thr Pro Ala Pro Val Arg Leu Asp Ala Gly Gly Pro Thr Val Pro
210 215 220

Met Arg Tyr Val Pro Tyr Ser Gly Leu Val Pro Thr Val Val Pro Asp
225 230 235 240

Trp Leu Arg Arg Pro Pro Glu Arg Pro Arg Val Leu Val Thr Leu Gly
245 250 255

Ile Thr Ser Arg Arg Val Lys Ser Phe Leu Ala Val Ser Val Asp Asp
260 265 270

Leu Phe Glu Ala Val Ala Gly Leu Gly Val Glu Val Val Ala Thr Leu
275 280 285

Asp Ala Asp Gln Arg Glu Leu Leu Gly Arg Val Pro Asp His Phe Arg
290 295 300

Ile Val Glu His Val Pro Leu Asp Ala Val Leu Pro Thr Cys Ser Ala
305 310 315 320

Ile Val His His Gly Gly Ala Gly Thr Trp Ser Thr Ala Ala Val Tyr
325 330 335

Gly Val Pro Gln Val Ser Leu Gly Ser Met Trp Asp His Phe Tyr Arg
340 345 350

Ala Arg Arg Leu Glu Glu Leu Gly Ala Gly Leu Arg Leu Pro Ser Gly
355 360 365

Glu Leu Thr Ala Glu Gly Leu Arg Thr Arg Leu Glu Arg Val Leu Gly
370 375 380

Glu Pro Ser Phe Gly Thr Ala Ala Gln Ala Leu Ser Asp Thr Ile Ala
385 390 395 400

Ala Glu Pro Ser Pro Ser Glu Val Val Pro Val Leu Glu Glu Leu Thr
405 410 415

Gly Arg His Arg Pro Gly Thr Arg
420

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 139 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snoL"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Met Ser Thr Thr Ala Asn Lys Glu Arg Cys Leu Glu Met Val Ala Ala
1 5 10 15

Trp Asn Arg Trp Asp Val Ser Gly Val Val Ala His Trp Ala Pro Asp
20 25 30

Val Val His Tyr Asp Asp Glu Asp Lys Pro Val Ser Ala Glu Glu Val
35 40 45

Val Arg Arg Met Asn Ser Ala Val Glu Ala Phe Pro Asp Leu Arg Leu
50 55 60

Asp Val Arg Ser Ile Val Gly Glu Gly Asp Arg Val Met Leu Arg Ile
65 70 75 80

Thr Cys Ser Ala Thr His Gln Gly Val Phe Met Gly Ile Ala Pro Thr
85 90 95

Gly Arg Lys Val Arg Trp Thr Tyr Leu Glu Glu Leu Arg Phe Ser Glu
100 105 110

Ala Gly Lys Val Val Glu His Trp Asp Val Phe Asn Phe Ser Pro Leu
115 120 125

Phe Arg Asp Leu Gly Val Val Pro Asp Gly Leu
130 135

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 155 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note="translate of snco"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

~~Met Ser Val Arg Thr Asp Gln Thr Ala Ala Pro Glu Asp Arg Ala Ala~~
1 5 10 15

Ala Thr Asp Pro Gly Phe Gly His Leu Tyr Ala Gln Val Gln Gln Phe
20 25 30

Tyr Ala Arg Gln Met Gln Leu Leu Asp Ser Gly Ala Ala Glu Glu Trp
35 40 45

Ala Ala Thr Phe Thr Glu Asp Gly Thr Phe Ala Arg Pro Ser Ser Pro
50 55 60

Glu Pro Ala Arg Gly His Ala Glu Leu Ala Ala Gly Ala Arg Ala Ala
65 70 75 80

Ala Glu Arg Leu Ala Ala Glu Gly Leu Ser His Arg His Val Ile Gly
85 90 95

Met Thr Ala Val Arg Arg Glu Pro Asp Gly Ser Val Phe Val Arg Ser
100 105 110

Tyr Ala Gln Val Phe Ala Thr Arg Arg Gly Glu Ala Pro Arg Leu His
115 120 125

Leu Ile Cys Val Cys Glu Asp Val Leu Val Arg Glu Gly Pro Gly Leu
130 135 140

Lys Val Arg Glu Arg Val Val Thr His Asp Ala
145 150 155

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 281 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snoaF"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Val Arg Ala Met Thr Asp Ser Thr Gly Pro Arg Pro Val Pro Ala Met
1 5 10 15

Ser Pro Ala Pro Ser Pro Thr Pro Ser Pro Gly Pro Ala Pro Gly Ser
20 25 30

Glu Pro Ala Pro Leu Ala Val Ile Val Thr Gly Gly Gly Ser Gly Ile
35 40 45

Gly Arg Ala Thr Ala Arg Ala Phe Ala Ala Gln Gly Ala Lys Val Leu
50 55 60

Val Val Gly Arg Thr Glu Asp Ala Leu Ala Gln Thr Ala Glu Gly Cys
65 70 75 80

Ala Asp Met Arg Val Leu Val Ala Asp Val Ala Ser Pro Asp Gly Pro
85 90 95

Gln Ala Val Val Asn Ala Ala Leu Arg Glu Phe Gly Arg Ile Asp Val
100 105 110

Leu Val Asn Asn Ala Ala Val Ala Gly Met Glu Thr Leu Gln Thr Val
115 120 125

Asp Arg Asp Ala Val Ala Arg Gln Phe Gly Thr Asn Leu Thr Ala Pro
130 135 140

Leu Phe Leu Val Gln Ser Ala Leu Gly Ala Leu Glu Lys Ser Arg Gly
145 150 155 160

Ile Val Val Asn Val Gly Thr Ala Ala Thr Leu Gly Leu Arg Ala Ala
165 170 175

Pro Thr Gly Ala Leu Tyr Gly Ala Ser Lys Val Ala Leu Asp Tyr Leu
180 185 190

Thr Arg Thr Trp Ala Val Glu Leu Ala Pro Arg Gly Ile Arg Val Val
195 200 205

Gly Val Ala Pro Gly Val Ile Asp Thr Gly Ile Gly Val Arg Met Gly
210 215 220

Met Thr Pro Glu Gly Tyr Arg Glu Phe Leu Thr Gly Met Gly Gly Arg
225 230 235 240

Val Pro Val Gly Arg Val Gly Arg Pro Glu Asp Val Ala Trp Trp Ile
245 250 255

Val Gln Leu Ala Arg Pro Glu Ala Gly Tyr Ala Thr Gly Met Val Val
260 265 270

Pro Val Asp Gly Gly Leu Ser Leu Val
275 280

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 190 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(D) OTHER INFORMATION: /note= "translate of snon"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

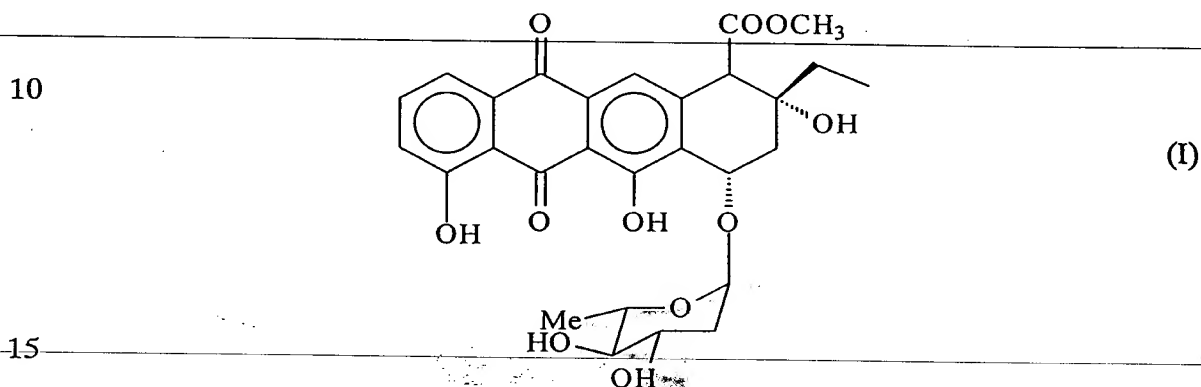
| | | | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Val | Gln | Glu | Thr | Glu | Pro | Gly | Val | Pro | Ala | Asp | Leu | Pro | Ala | Glu | Ser | 1 | 5 | 10 | 15 |
| Asp | Pro | Ala | Ala | Leu | Glu | Arg | Leu | Ala | Ala | Arg | Tyr | Arg | Arg | Asp | Gly | 20 | 25 | 30 | |
| Tyr | Val | His | Val | Pro | Gly | Val | Leu | Asp | Ala | Gly | Glu | Val | Ala | Glu | Tyr | 35 | 40 | 45 | |
| Leu | Ala | Glu | Ala | Arg | Arg | Leu | Leu | Ala | His | Glu | Glu | Ser | Val | Arg | Trp | 50 | 55 | 60 | |
| Gly | Ser | Gly | Ala | Gly | Thr | Val | Met | Asp | Tyr | Val | Ala | Asp | Ala | Gln | Leu | 65 | 70 | 75 | 80 |
| Gly | Ser | Asp | Thr | Met | Arg | Arg | Leu | Ala | Thr | His | Pro | Arg | Ile | Ala | Ala | 85 | 90 | 95 | |
| Leu | Ala | Glu | Tyr | Leu | Ala | Gly | Ser | Pro | Leu | Arg | Leu | Phe | Lys | Leu | Glu | 100 | 105 | 110 | |
| Val | Leu | Leu | Lys | Glu | Asn | Lys | Glu | Lys | Asp | Ala | Ser | Val | Pro | Thr | Ala | 115 | 120 | 125 | |
| Pro | His | His | Asp | Ala | Phe | Ala | Phe | Pro | Phe | Ser | Thr | Ala | Gly | Thr | Ala | 130 | 135 | 140 | |
| Leu | Thr | Ala | Trp | Val | Ala | Leu | Val | Asp | Val | Pro | Val | Glu | Arg | Gly | Cys | 145 | 150 | 155 | 160 |
| Met | Thr | Phe | Val | Pro | Gly | Ser | His | Leu | Leu | Pro | Asp | Pro | Asp | Thr | Gly | 165 | 170 | 175 | |
| Asp | Glu | Pro | Trp | Ala | Gly | Ala | Phe | Thr | Arg | Pro | Gly | Glu | Ile | 180 | 185 | 190 | | | |

Claims

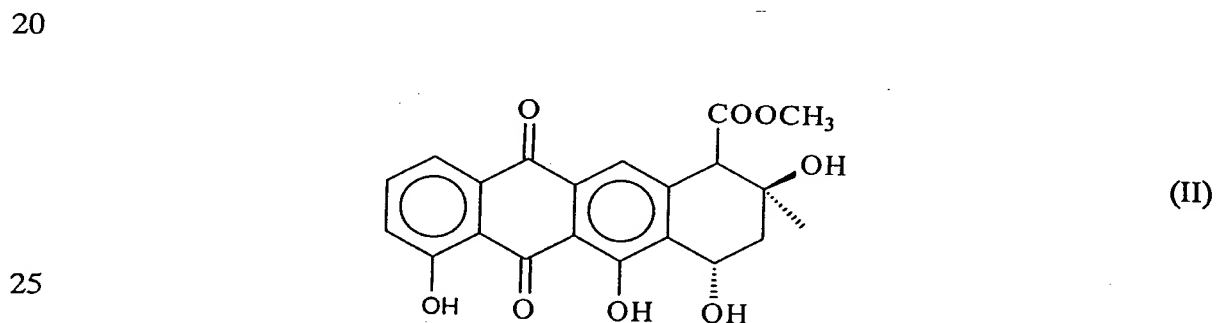
1. Isolated and purified DNA fragment, which is the gene cluster for the anthracycline biosynthetic pathway of the bacterium *Streptomyces nogalater*, being included in a 10kb and a 7kb flanked *Bgl*II fragments of *S. nogalater* genome.
2. ~~The DNA fragment according to claim 1, comprising the nucleotide sequence given in SEQ ID NO:1, or a sequence showing at least 80% homology to said sequence.~~
3. A recombinant DNA, which comprises the DNA fragment according to claim 1 or 2, cloned in a plasmid replicating in *Streptomyces*.
4. The recombinant DNA according to claim 3, which is the plasmid pSY15c, comprising a 1.4 kb *Bam*HI-*Sac*I fragment from the plasmid pSY42 and a 1.1 kb *Mlu*I-*Kpn*I fragment from the plasmid pSY43.
5. Plasmid pSY42, deposited in *S. lividans* strain TK24/pSY42 with the deposition number DSM 12451.
6. Plasmid pSY43, deposited in *S. lividans* strain TK24/pSY43 with the deposition number DSM 12452.
7. A process for the production of hybrid compounds, comprising transferring the DNA fragment according to claim 1 or 2 into a *Streptomyces* host, cultivating the recombinant strain obtained, and isolating the compounds produced.
8. The process according to claim 7, wherein the *Streptomyces* host is a *Streptomyces galilaeus* host.

9. The process according to claim 8, wherein the *Streptomyces galilaeus* host is selected from the strains H026, H039, H063 and H075, which are mutant strains of *S. galilaeus* ATCC 31615.

10. The process according to claim 8, wherein an anthracycline is produced, which has the following formula I



11. The process according to claim 8, wherein an anthracyclinone is produced, which has the following formula II



30 12. A process for the production of hybrid compounds, comprising transferring at least one of the genes selected from the group consisting of *snogJ*, *snogA*, *snoaM*, *snogN*, *snoaG*, *snogC*, *snogK*, *snoaL*, *snoK*, *snogD*, *snoW*, *snogE*, *snoL*, *snoO* and

snoaF into a *Streptomyces* host, said genes being derived from the DNA fragment of claim 1 or 2, cultivating the recombinant strain obtained, and isolating the compounds produced.

- 5 13. The process according to claim 12, wherein the gene *snoaL* encoding NAME cyclase is transferred into a *Streptomyces* host.

14. The process according to claim 12, wherein at least one of the genes *snoG*D and *snoG*E encoding glycosyl transferases is transferred into a *Streptomyces* host.

10

15. The process according to claim 12, wherein at least one of the genes *snoG*J, *snoG*N, *snoG*C, *snoG*K and *snoG*A affecting the formation of nogalamine and nogalose is transferred into a *Streptomyces* host.

Abstract

5

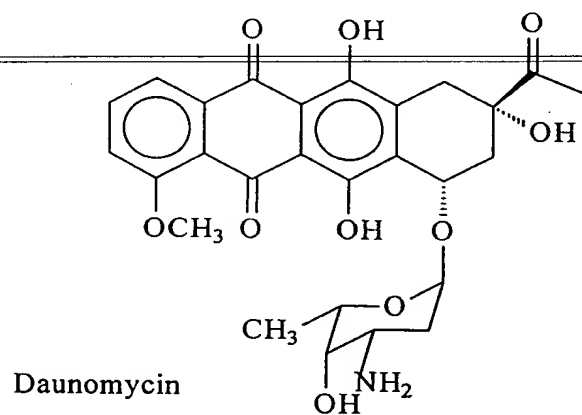
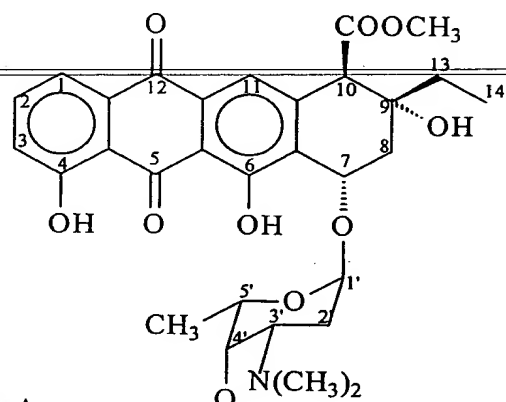
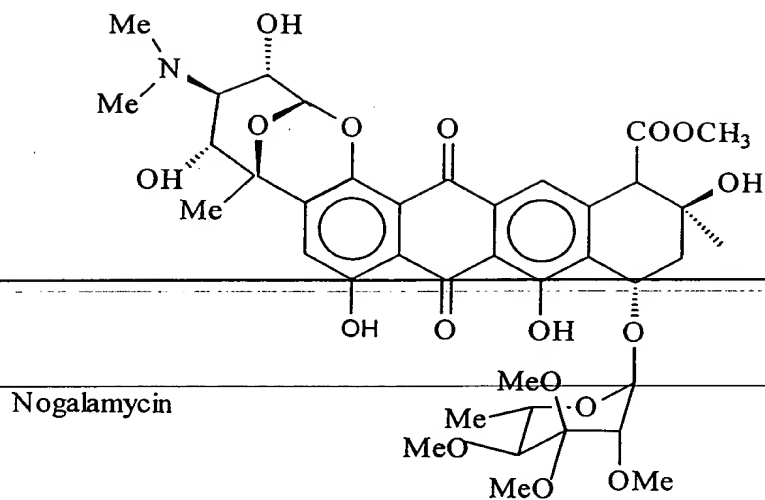
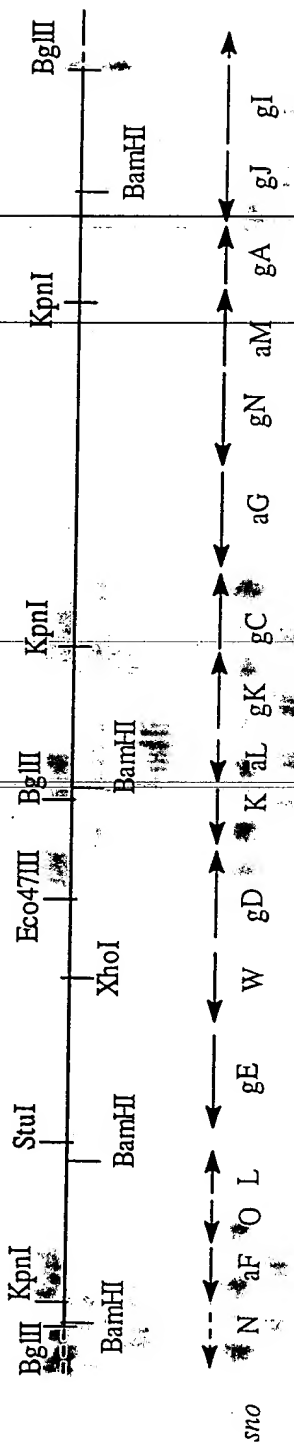


Fig. 1

1kb



SY43

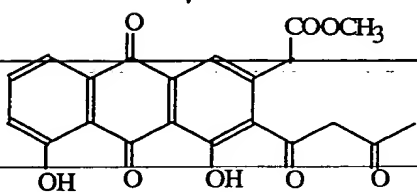
SY42

Fig. 2

Aglycone moiety pathway

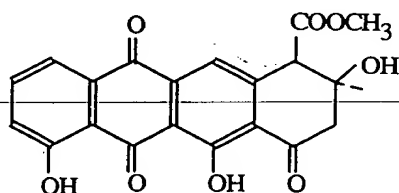
Sugar residue pathway

10 Acetates



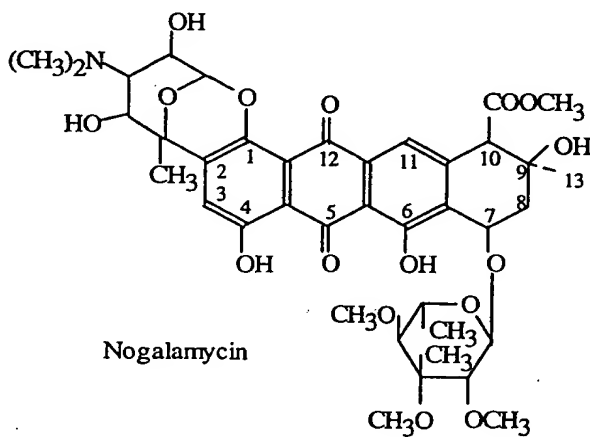
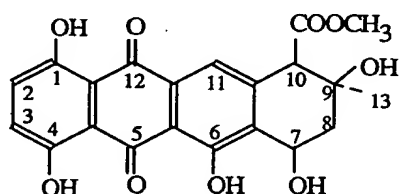
NAME

NAMEC (*snoaL*)

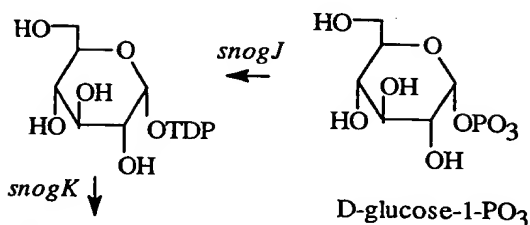


C-7 KR (*snoaF*)

snoaG

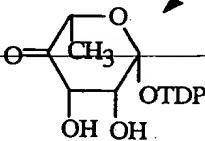
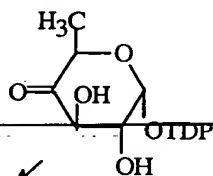


Nogalamycin

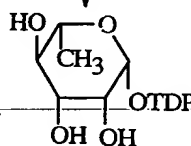


D-glucose-1-PO₃

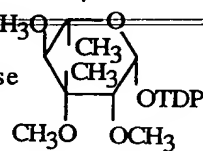
snogK



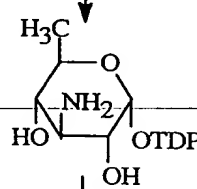
snogC



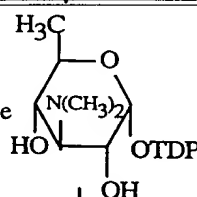
Nogalose



snogI



snogA



Nogalamine

snogE

snogD

Fig. 3

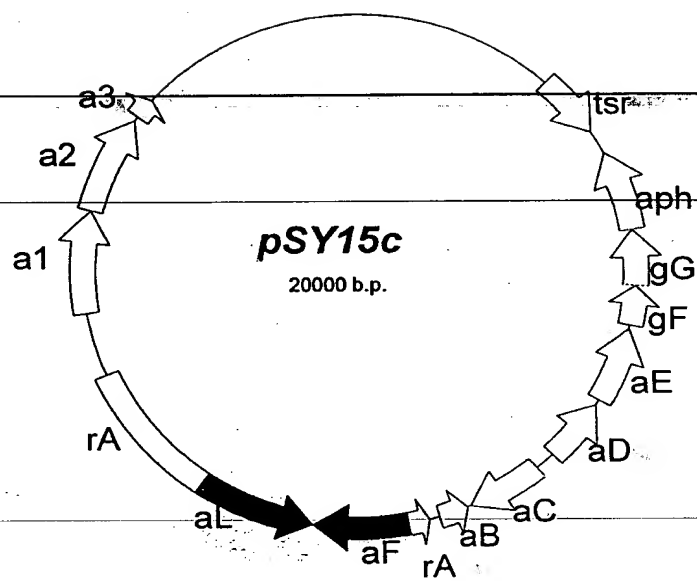


Fig. 4